

14 0 7113

Journal of
**Geophysical
Research**

VOLUME 65

JANUARY 1960

NUMBER 1

**THE SCIENTIFIC PUBLICATION
OF THE AMERICAN GEOPHYSICAL UNION**

Journal of Geophysical Research

An International Scientific Publication

OFFICERS OF THE UNION

LLOYD V. BERKNER, *President*
F. W. REICHELDERFER, *Vice President*
A. NELSON SAYRE, *General Secretary*
WALDO E. SMITH, *Executive Secretary*

OFFICERS OF THE SECTIONS

Geodesy

CHARLES PIERCE, *President*
FLOYD W. HOUGH, *Vice President*
BUFORD K. MEADE, *Secretary*

Seismology

LEONARD M. MURPHY, *President*
JAMES A. PEOPLES, JR., *Vice President*
BENJAMIN F. HOWELL, JR., *Secretary*

Meteorology

THOMAS F. MALONE, *President*
GORDON E. DUNN, *Vice President*
WOODROW C. JACOBS, *Secretary*

Geomagnetism and Aeronomy

L. R. ALDREDGE, *President*
C. T. ELVEY, *Vice President*
J. HUGH NELSON, *Secretary*

Oceanography

WALTER H. MUNK, *President*
DONALD W. PRITCHARD, *Vice President*
EUGENE C. LAFOND, *Secretary*

Volcanology, Geochemistry, and Petrology

ALFRED O. C. NIER, *President*
FRANCIS J. TURNER, *Vice President*
IRVING FRIEDMAN, *Secretary*

Hydrology

WALTER B. LANGBEIN, *President*
WILLIAM C. ACKERMANN, *Vice President*
CHARLES C. McDONALD, *Secretary*

Tectonophysics

PATRICK M. HURLEY, *President*
LOUIS B. SLICHTER, *Vice President*
H. RICHARD GAULT, *Secretary*

BOARD OF EDITORS

Editors: PHILIP H. ABELSON and J. A. PEOPLES, JR.

ASSOCIATE EDITORS

1960

HENRY G. BOOKER	WALTER B. LANGBEIN
E. C. BULLARD	ERWIN SCHMID
JULE CHARNEY	HENRY STOMMEL
GEORGE T. FAUST	J. TH. THIJSSÉ
DAVID G. KNAPP	A. H. WAYNICK

J. TUZO WILSON

1960-1961

HENRI BADER	T. NAGATA
K. E. BULLEN	FRANK PRESS
CONRAD P. MOOK	A. NELSON SAYRE
WALTER H. MUNK	MERLE A. TUVE

JAMES A. VAN ALLEN

1960-1962

JULIUS BARTELS	TOR J. NORDENSON
E. G. BOWEN	E. N. PARKER
JOHN E. CHAPPELEAR	GEORGE P. RIGSBY
G. D. GARLAND	WALTER O. ROBERTS
GORDON J. F. MACDONALD	C. N. TOUART
L. A. MANNING	JAMES R. WAIT

This Journal welcomes original scientific contributions on the physics of the earth and its environment.

Manuscripts should be transmitted to J. A. Peoples, Jr., Geology Department, University of Kansas, Lawrence, Kansas. Authors' institutions, if in the United States or Canada, are requested to pay a publication charge of \$15 per page, which, if honored, entitles them to 100 free reprints.

Subscriptions to the *Journal of Geophysical Research and Transactions, AGU* are included in membership dues.

Non-member subscriptions, *Journal of Geophysical Research*...\$30 for back Volume of 1959, \$5 for this issue; \$20 for the calendar year 1960.

Non-member subscriptions, *Transactions, AGU*...

.....\$4 per calendar year, \$1.25 per copy.

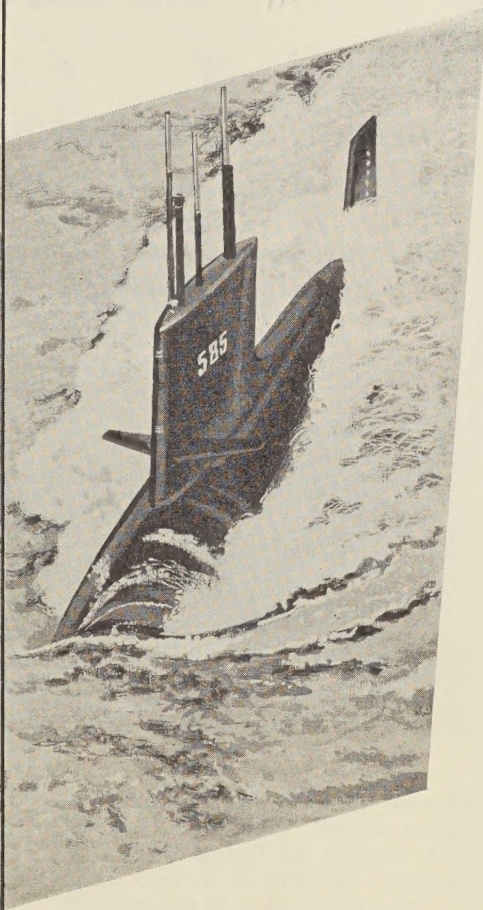
Subscriptions, renewals, and orders for back numbers should be addressed to American Geophysical Union, 1515 Massachusetts Ave., Northwest, Washington 5, D. C. Suggestions to authors are available on request.

Advertising Representative: Howland and Howland, Inc., 230 Park Ave., New York 17, N. Y.

Beginning with the January 1959 issue (Vol. 64, No. 1) the *Journal of Geophysical Research* is published monthly by the American Geophysical Union, the U. S. National Committee of the International Union of Geodesy and Geophysics organized under the National Academy of Sciences-National Research Council, as the U. S. national adhering body. Publication of this journal is supported by the National Science Foundation and the Carnegie Institution of Washington. The new monthly combines the type of scientific material formerly published in the bi-monthly *Transactions, American Geophysical Union*, and the quarterly *Journal of Geophysical Research*. The *Transactions, American Geophysical Union* will continue as a quarterly publication for Union business and items of interest to members of the Union.

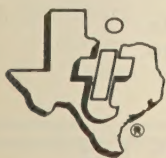
Published monthly by the American Geophysical Union from 1407 Sherwood Avenue, Richmond, Virginia. Second class postage paid at Richmond, Virginia.

from seismic detection...



to ASW

For information on the broad capabilities of Texas Instruments in fundamental and applied earth sciences — serving defense requirements and the mineral industries — write for Bulletin DM-101.

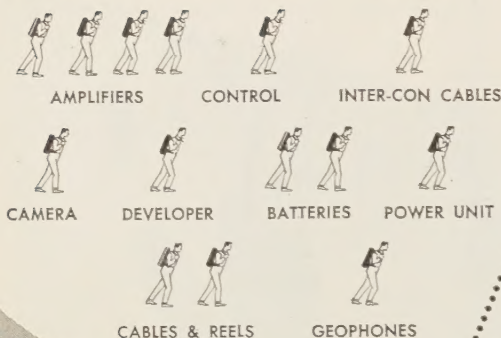


GEOSCIENCES & INSTRUMENTATION DIVISION
TEXAS INSTRUMENTS
INCORPORATED

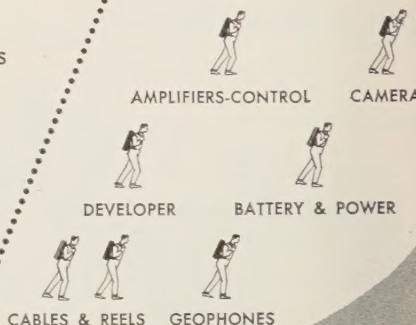
P. O. Box 35084, Airlawn Station, Dallas 35, Texas

Please mention JOURNAL OF GEOPHYSICAL RESEARCH, when writing to advertisers

**TYPICAL FULL RANGE
24-trace seismograph**



**EXPLORER
FULL RANGE
24-trace seismograph**



**Cut manpower in half
...double production with**

EXPLORER

ALL-TRANSISTORIZED SEISMOGRAPH

In some of the world's toughest assignments, the EXPLORER Model 8000 24-trace All-Transistorized Seismograph has effected reductions in recording personnel of 50 percent from the number required for the typical full range portable systems. Due to the reduced weight and number of packages for amplifiers, batteries and interconnecting cables, only seven men are required for the 8000 system. In foreign operations, the additional savings in manpower will be several times the number of packages eliminated due to release of support personnel.

Of most importance, it is reported that

the decreased time involved at and between each recording set-up with the EXPLORER Seismograph has approximately doubled production, and, of course, cut the cost per profile in half.

Furthermore, the 8000 system is effective over a wide range of frequencies, from high resolution to refraction. You have a wide selection of AGC fixed and TVG gain controls.

The EXPLORER is the only completely *field-proved* All-Transistorized Seismograph System. Let the TI engineer show you how it pays for itself in *reduced operating costs, increased production, and unequalled reliability.*

***Now operating in Sumatra, Peru, Columbia,
Mexico, France, Canada and South Louisiana.**

Write for Bulletin No. S-324



**TEXAS INSTRUMENTS
INCORPORATED**

GEOSCIENCES & INSTRUMENTATION DIVISION
3609 BUFFALO SPEEDWAY • HOUSTON 6, TEXAS • CABLE: TEXINS

Lockheed Missiles and Space Division

Offers

PHYSICISTS

Unusual Opportunities in Research and Development

■ Important career positions are available at Lockheed Missiles and Space Division at its new facilities on the beautiful San Francisco Peninsula—one of the choicest living areas in the nation. Headquarters for the Division are at Sunnyvale, California, with Research and Development facilities located in the Stanford Industrial Park in nearby Palo Alto. Equipment and facilities are completely modern and include one of the largest computer centers in the world.

Interesting opportunities exist for physicists to contribute to the solution of new problems in the fields listed below. Please write: Research and Development Staff, Dept. A-59, 962 West El Camino Real, Sunnyvale, Calif. U.S. citizenship or existing Department of Defense security clearance required.

SPACE PHYSICS Advanced degree preferred, for work in basic research on the physics of the earth's upper atmosphere and beyond. Typical projects include: measurement of atmospheric composition and density at satellite altitudes; laboratory experiments on upper atmospheric atomic and molecular reactions; hydromagnetic interactions with the earth's magnetic fields; simulation and study of meteor impacts; and particle radiation.

INFRARED AND OPTICS Advanced degree in E.E. or physics preferred, with experience in infrared systems research and development, or electrical engineers with background in electronics information theory; servo-mechanisms; specialized circuitry, as in low-level voltage circuits, or physicists with background in optics or semiconductors. For work in infrared physics research; advanced systems development; or physical measurements in infrared. Research is also being conducted in optical devices and systems, including scanners, encoders, detectors, and read-out devices.

SOLID STATE DEVICES Advanced degree required, Ph.D. preferred, in E.E. or physics and evidence of creative, original work through published articles, patents or superior Ph.D. theses for research work in one or more of the following: thermoelectric; photovoltaic; lumistor; ferrite; logic component; sensor; thermistor; or cryogenic devices. Also, materials analysis and evaluation; processing techniques design and development of novel electronics devices and components; circuit analysis; circuit topology; or microminiaturization.

***Lockheed* / MISSILES AND SPACE DIVISION**

Systems Manager for the Navy POLARIS FBM; the Air Force AGENA Satellite in the DISCOVERER Program; MIDAS and SAMOS; Air Force X-7; and Army KINGFISHER.

SUNNYVALE, PALO ALTO, VAN NUYS, SANTA CRUZ, SANTA MARIA, CALIFORNIA

CAPE CANAVERAL, FLORIDA • ALAMOGORDO, NEW MEXICO • HAWAII

THE PHYSICS OF FLUIDS

A JOURNAL PUBLISHED BY THE AMERICAN INSTITUTE OF PHYSICS

Forthcoming Papers

Stability of an Idealized Atmosphere	K. M. Case
N-Wave Propagating into a Stratified Atmosphere	S. G. Reed, J. K. Loquicini
Flow and Stress Near an Interface between Stratified Liquids	K. Loquicini
Stability of Inviscid Plane Couette Flow	K. M. Case
Cross-Stresses in Laminar Flow of Liquids	M. Reinhold
Heat-Transport by Convection	Y. Nakagawa
Heat-Transport by Convection in Presence of an Impressed Magnetic Field	Y. Nakagawa
One-Dimensional Equilibrium Spectra in Isotropic Turbulence	W. H. Rees
A Theory of Decaying Homogeneous Turbulence	Robert G. Deissler
Growth of Instabilities on Displacement Fronts in Porous Media ..	A. E. Scheidegger
A Kinetic Theory of Liquid Displacement	A. J. de Wit
Boundary Layer over a Flat Plate in Presence of Shear Flow	Lu Tien
Hypersonic Flow Near Leading Edge of a Flat Plate ..	H. T. Nagamatsu and T. Y. Li
Absorption Coefficients for High-Temperature Nitrogen, Oxygen, and Air	B. H. Armstrong and R. E. Meyerowitz
Wake of Satellite Traversing the Ionosphere	S. Rast
Experimental Study of Alfvén-Wave Properties	J. M. Wilcox, F. I. Boley, and A. W. DeSilva
Test Particles in a Completely Ionized Plasma	Norman Rostoker and M. N. Rosenbluth
Hydromagnetic Stability of Streaming Cylindrical Incompressible Plasma	N. J. Zabusky
Instability of Contra-Streaming Plasmas	Paul J. Kellogg and Harold Liemohn
Performance of a Hydromagnetic Plasma Gun	John Marshall
Absorption of Plasma Waves	Thomas H. Stix
On the Electrical Behavior of an Ideal Plasma	Giorgio Gambirasio
Transverse Compression Waves in a Stabilized Discharge	Daryl Reagor
Cyclotron Radiation from a Hot Plasma ...	W. E. Drummond and M. N. Rosenbluth
Flow of a Partially Ionized Gas in an Axial Magnetic Field	P. J. Dickerman and C. F. Prior
Ion Wave Instabilities	I. B. Bernstein, E. A. Frieman, R. M. Kulsrud and M. N. Rosenbluth
Irreversible Processes in Ionized Gases	R. Bales
Diagram Expansions in Quantum Statistics	Howard B. Levi
Asymptotic Value of the Pair Distribution Near a Wall	Joel L. Lebowitz
Remark on Dilute Bose Systems	J. Peregrine
Relaxation Theory of Thermal Conduction in Liquids	R. E. Nettleton
Isotope Separation by a Hot-Wire Thermal Diffusion Column	S. C. Saxena and W. W. Watson

Annual Subscription Rates: American Physical Society in lieu subscription—\$4 domestic, \$5 foreign;
Members of A.I.P. societies—\$8 domestic, \$9 foreign; Non-Members—\$10 domestic, \$11 foreign

Orders and inquiries should be sent to

AMERICAN INSTITUTE OF PHYSICS, 335 East 45th Street, New York 17, N. Y.

Please mention JOURNAL OF GEOPHYSICAL RESEARCH, when writing to advertisers

ASKANIA - GRAVIMETER

Type Gs 11 with Recording Equipment

The Gravity Meter most used for earth tidal measurements

Type Gs 12

The extremely reliable instrument with built-in ball system

Sea Gravimeter Gss 2 after Graf

with direct recording for uninterrupted measurements on board ship



ASKANIA - WERKE • AKTIENGESELLSCHAFT

U. S. Branch Office, 4913 Cordell Ave., Bethesda, Maryland

CAREER OPPORTUNITIES

Solving the Puzzles of **METEOROLOGY**

Allied Research scientists, engineers and technicians in the Geophysics Division, are working directly on the frontier of knowledge utilizing the latest scientific and engineering technological advances to perform studies in . . . satellite meteorology . . . radar meteorology . . . short-range weather forecasting . . . cloud and fog dispersal . . . and other areas as diverse as the weather itself.

The unusually broad capabilities of Allied Research have created outstanding opportunities for those dedicated to working on the frontier of knowledge in:

Meteorology • Geophysics • Physics Research • Nuclear Weapons Effects • Weapons Systems Analysis • Aerodynamics • Applied Mechanics • Electronics and Instrumentation • Systems Engineering • Vibration Isolation

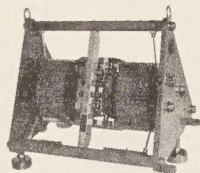


Discover the challenge and sense of achievement that only a compact, research-directed company can offer the professional man. Send resume, in confidence, to: Mr. Norman Metzger

ALLIED RESEARCH ASSOCIATES, Inc.

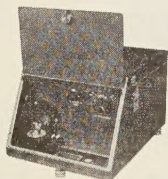
43 Leon Street, Boston, Massachusetts • GArrison 7-2434

RESEARCH • ENGINEERING • DEVELOPMENT



BENIOFF SEISMOMETERS HORIZONTAL and VERTICAL

- ▶ **SENSITIVE**—can detect 1 millimicron of earth motion at 1 cps; 8 separate coils; 1 mv output per coil per micron of motion at 1 cps.
- ▶ **DUAL PURPOSE**—can drive long and short-period galvanometers simultaneously.
- ▶ **POWERFUL**—magnification of 300,000 at 1 cps; can drive galvanometers remoted 10 miles without amplification.
- ▶ **CALIBRATION**—locally, or remote with Remote Calibration Kits 5227 or 5528.
- ▶ **RUGGED**—push-pull variable-reluctance transducer; large air gaps; readily adjusted; weather resistant plating; designed for years of trouble-free operation.



AMPLIFIER MODEL 4300

- ▶ **VOLTAGE GAIN**— $\frac{1}{2}$ million at 2 cps.
- ▶ **NOISE LEVEL**— 1×10^{-8} volts, p-p referred to input; employs .01 cps high-pass filter for long-term drift elimination; 70 db dynamic range.
- ▶ **DRIFT**—.1 microvolts in 8 hours; can operate unattended for months; separate power supply.
- ▶ **RUGGED**—heavy cast-aluminum construction; can be shipped anywhere.
- ▶ **PRINCIPAL**—twin photocells and a rugged Geotech Galvanometer. Series 4100, whose sensitivity is 2×10^{-8} amps/mm at 1 meter at 2 cps.
- ▶ **DELIVERY 30 DAYS**—write for information to:

THE GEOTECHNICAL CORP.

3401 Shiloh Road
Garland, Texas

P. O. Box 28277
Dallas 28, Texas



NEW REPRINT

American Geophysical Union: Transactions

(Reproduced with the permission of the American Geophysical Union)

Now Available

Volumes 13-15, 1932-1934

Volume 13, 1932, paper bound
Volume 14, 1933, paper bound
Volume 15, 1934, paper bound

Previously Reprinted

Volumes 1-12, 1920-1931

(Volumes 3 and 5 were never published)

Paper bound set (in 9 volumes)	\$110.00
Volume 1, 1920, paper bound	5.00
Volume 2, 1921, paper bound	10.00
Volume 4, 1923, paper bound	15.00
Volume 6, 1925, paper bound	5.00
Volume 7, 1926, paper bound	15.00
Volume 8, 1927, paper bound	20.00
Volume 9, 1928, paper bound	15.00
Volume 10-11, 1929-1930, paper bound	20.00
Volume 12, 1931, paper bound	15.00

(Volumes 2, 4, and 6-9 published in National Research Council Bulletin)

Volumes 16-34, 1935-1953, will be reproduced by photo-offset as soon as there is sufficient demand to warrant the undertaking of a reprint edition.



JOHNSON

REPRINT CORPORATION

111 FIFTH AVENUE

NEW YORK 3, NEW YORK

Journal of GEOPHYSICAL RESEARCH

VOLUME 65

JANUARY, 1960

No. 1

Charge and Magnetic Field Interaction with Satellites

DAVID B. BEARD

*Lockheed Missiles and Space Division, Palo Alto, Calif.
and the University of California, Davis.*

AND

FRANCIS S. JOHNSON

Lockheed Missiles and Space Division, Palo Alto, Calif.

Abstract. The interaction of a satellite with the magnetic field of the earth and the ionized medium through which it is moving has been investigated. Owing to the differing incident velocities of ions and electrons and therefore differing incident flux intensities, a negative potential will be induced on the satellite, but it is smaller than has previously been believed. Satellite motion across the magnetic lines of the earth will induce a voltage on the satellite of as much as 0.2 volt per meter of satellite size, and this may affect the interpretation of measurements of satellite potential. The magnetic drag resulting from the induced currents is proportional to the cube of the satellite dimensions and may exceed the mass drag for satellites larger than 50 meters in diameter; this can occur only above 1200-km altitude, where the charge density exceeds the neutral density. Thus the magnetically induced current is an insignificant cause of drag. Although some useful power can be extracted from the induced current, it is not a very promising source of auxiliary power for presently conceived vehicles.

I. Introduction. A satellite moving in a conducting fluid within a magnetic field will develop electric potentials relative to the surrounding medium which slightly alter its drag. For example, a metallic satellite moving across the earth's magnetic field will develop a potential per unit length perpendicular to the field of $10^{-8} HV$ volts/cm, where H is the magnetic field in gauss and V the satellite velocity in centimeters per second. Although this induced voltage has negligible effect on the ion motion, it does influence the incident electron current as a function of position on the satellite surface, causing the electron current to occur entirely at the high-potential end of a large vehicle. The resulting electric current within the satellite will interact with the magnetic field, causing a magnetic drag in addition to the generally discussed mass drag. Under favorable circum-

stances, such as satellite motion perpendicular to the terrestrial magnetic field near altitudes where the electron concentration is maximum, the magnetically induced emf may result in a small useful source of auxiliary power to charge batteries or operate a radio transmitter.

In addition to this magnetically induced voltage, there exists an over-all satellite potential which has been discussed by *Jastrow and Pearse* [1957]. (The calculation of the average satellite potential presented here differs from that by Jastrow and Pearse in several ways, three of which are: (1) the effect of the magnetic field is considered in this paper; (2) kT is assumed to be about 0.1 ev instead of 1.5 ev; and (3) the flux of incident electrons was computed exactly rather than approximately). The ambient ions and electrons move at different fluxes of positive and negative charges to a surface at ground

potential. The electrons have the higher velocities and cause an over-all negative satellite potential to result in order to maintain an over-all zero electric current into the satellite surface.

The salient features of the medium are discussed in section II. The resulting surface potentials and incident current densities are calculated as a function of position on the surface and in toto in section III. The magnetic drag and suggested useful power possibilities are presented in section IV. Drift currents and other small charge effects are mentioned briefly in section V, and the general consequences of considering the satellite motion in an ionized medium within a magnetic field are presented in the concluding section VI.

II. *Pertinent numerical quantities for charge behavior.* For an examination of the charge interaction with the satellite surface, it is well to first list some pertinent velocities and distances. kT is assumed to be about 0.1 ev (i.e., $T \approx 1000^\circ\text{K}$), and the particle density, $N_{\text{ions}} = N_{\text{electrons}}$, is assumed to be $10^6/\text{cc}$, which is a maximum concentration occurring at about 400-km altitude. This decreases, for example, to about $5 \times 10^4/\text{cc}$ at 1200-km altitude.

Cyclotron period for electrons:

$$\tau_c = 2\pi m_e c / eH \sim 6 \times 10^{-7} \text{ sec.}$$

Cyclotron radius for electrons:

$$a_e = m_e VC / eH \sim 0.07 \sqrt{T} \sim 2 \text{ cm.}$$

Cyclotron radius for ions:

$$a_i \sim 340 \text{ cm.}$$

Debye shielding distance:

$$\begin{aligned} \lambda_D &= \sqrt{kT / 4\pi N_e e^2} \\ &= 6.9 \sqrt{T / N_e} \sim 0.2 \text{ cm,} \end{aligned}$$

which is about 10 per cent of the cyclotron radius for electrons, i.e.,

$$\lambda_D / a_e \sim 100 / \sqrt{N_e} \sim 10 \text{ per cent.}$$

Average electron velocity:

$$\bar{v}_e \sim 2 \times 10^7 \text{ cm/sec.}$$

Average ion velocity:

$$\bar{v}_i \sim 10^5 \text{ cm/sec.}$$

Satellite velocity:

$$V \sim 8 \times 10^5 \text{ cm/sec}$$

$$\sim 0.5 \text{ cm/electron cyclotron period}$$

Collision frequency for electrons:

$$\begin{aligned} \bar{\nu}_e &\sim 3.7 \times 10^{-9} n / E^{3/2} (\text{kev}) / \text{sec} \\ &\sim 3.7 \times 10^3 / \text{sec.} \end{aligned}$$

Mean free path for ions and electrons:

$$\lambda_{ee} \sim 5.4 \times 10^3 \text{ cm.}$$

The collision frequency and mean free path were computed on the basis of collisions between charged particles only. Although charged particles interact for distances up to a Debye length, which is roughly 10^3 times as large as atomic dimensions, collisions with neutrals become the dominant cause of ion deflection below about 250-km altitude. The reasons are that (1) the interaction range only enters in the logarithm of the expected deflection angle for charged particle interactions, (2) the distance of closest approach upon which the deflection angle depends quadratically is about 100 times as large as atomic dimensions for charged particles at this temperature, and (3) neutrals are more than 10^4 times as abundant as charged particles below 250-km altitude.

The electron velocity distribution is surely Maxwellian at night, but during the daytime there must be a small bulge on the high-velocity tail of the distribution, due to energetic photoelectrons that have not yet slowed down (Jastrzebski and Pearse, 1957). This non-Maxwellian bulge is neglected in our calculations on the basis that it produces only a negligible effect on the charge attained by a satellite or rocket. The experimental evidence in support of this position is that rocket-borne mass-spectrometer analyses show the ion composition in the ionosphere has not changed, shown that no charges larger than a few tenths of a volt accumulate on rockets unless there are powerful transmitters in the rocket (C. Johnson, private communication). The few tenths of a volt can be explained in terms of a pure Maxwellian velocity distribution, and the large potentials on rockets or satellites with powerful transmitters can be explained in terms of the interaction between the RF fields and the electron

around the vehicle. If the non-Maxwellian age on the tail of the velocity distribution is significant in this connection, rockets with γ -powered transmitters would develop larger potentials than the observed few tenths of a volt.

II. Satellite surface potentials and current densities. The satellite will run into the ions, is acquiring a positive charge at the rate of $7l_x l_y$ coulombs/sec, where the y axis is in the direction of the satellite motion and $l_x l_y$ is taken to be the satellite area perpendicular to the satellite velocity. As previously discussed by Jastrow and Pearse [1957], this positive current must be opposed by an equal current of faster electrons. The mean motion of the electrons must be parallel to the magnetic field. The electron current is controlled by the potential of the satellite surface, which is

$$\phi = \phi_0 + 10^{-8} H V x \quad (1)$$

where ϕ_0 is the satellite surface potential at $x = 0$, the surface midpoint along the x direction. Assume a rectangular parallelepiped satellite having perpendicular to the magnetic field and parallel to one of its axes, and $H = H_z$ (see Fig. 1).

$$e l_y V = 2 n e l_y \sqrt{\frac{m_e}{2 \pi k T}} \left\{ \int_a^{l_x/2} dx \int_0^\infty v_x dv_x e^{-m_e v_x^2 / 2 k T} - \int_{-l_x/2}^a dx \int_{\sqrt{(2/m_e)(-e\phi)}}^\infty v_x dv_x e^{-m_e v_x^2 / 2 k T} \right\} \quad (2)$$

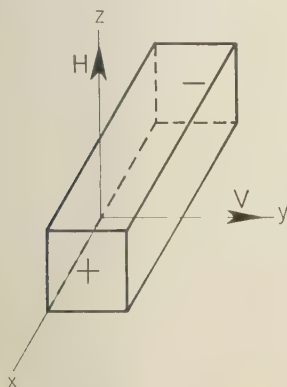


Fig. 1. Schematic representation of the induced charge on a satellite moving perpendicular to the magnetic field.

where a , the point at zero potential, is defined by

$$\phi_0 = -10^{-8} H V a \quad (3)$$

If $a > l_x/2$, only the second integral on the right-hand side of equation 2 is appropriate, with $l_x/2$ substituted for a . Equations 2 and 3 serve to define ϕ_0 . If the satellite is too small for the magnetically induced voltage to exceed the average satellite potential, that is $l_x/2 \leq a$, then

$$\begin{aligned} \frac{l_x l_y}{l_y} V &= \sqrt{\frac{2kT}{\pi m_e}} \frac{2kT}{e 10^{-8} H V} \\ &\cdot e^{e\phi_0/kT} \sinh \frac{e 10^{-8} H V l_x}{2kT} \\ &= -\sqrt{\frac{2kT}{\pi m_e}} \frac{2kT}{e\phi_0} a e^{e\phi_0/kT} \sinh -\frac{e\phi_0}{kT} \frac{l_x/2}{a} \quad (4) \\ &\approx \sqrt{\frac{2kT}{\pi m_e}} l_x e^{e\phi_0/kT} \quad \text{for } l_x/2 \ll a \end{aligned}$$

This last approximation yields

$$\frac{e\phi_0}{kT} = -\log_e \sqrt{\frac{2kT}{\pi m_e}} \frac{l_y}{l_x} \frac{1}{V} \quad (5)$$

the electrical potential equation for a satellite which is in a plasma with no magnetic field or whose motion is parallel to the magnetic field.

In the special case $l_y = l_x$, $e\phi_0/kT = 2.6$ at this assumed temperature of 0.1 ev. That is, the satellite will be charged to a negative potential of about a quarter of a volt. An excess of positive ions will surround the satellite for a distance, given correctly by Jastrow and Pearse, of

$$l = \sqrt{2 |e\phi| / kT \lambda_D}$$

resulting in an electric potential of

$$\phi = \phi_0 [1 - s/l]^2$$

at the distance s from the surface. $\phi = 0$ for $s > l$.

The condition for the validity of equation 4 is that $l_x/2 \leq a$. Letting $e\phi_0/kT = \alpha$, $l_x/2 = a$, and setting

$$2 \sinh(-\alpha) = e^{-\alpha}$$

the following condition for the validity condition is obtained from equation 4:

$$(l_x/l_y)(V/\bar{v}_e) = 1/4\alpha$$

If $l_z = l_y$ and $kT = 0.1$ ev,

$$\alpha = \bar{v}_e/4V = 6.6$$

Then

$$l_z = 13 kT/e 10^{-8} HV = 540 \text{ cm} \quad (6)$$

in a magnetic field of 0.3 gauss. For a satellite of this dimension, $\phi_0 \sim -0.7$ volt, and one end of the satellite is at a potential of -1.3 volts. This, incidentally, is one of several effects, such as the electrons absorbing energy from electromagnetic radiation from the antennas as suggested by Wm. B. Hanson (private communication), which could contribute to the anomalously high satellite potentials recently reported by the Russians.

For satellites larger than about 10 meters, specifically if $l_z/2 > a$,

$$\frac{l_z l_y}{l_y} V = \sqrt{\frac{2kT}{\pi m_e}} \left\{ \frac{l_z}{2} - a \left(1 + \frac{kT}{e\phi_0} \right) + \frac{kT}{e\phi_0} a \exp \left[\frac{e\phi_0}{kT} \left(1 + \frac{l_z/2}{a} \right) \right] \right\} \quad (7)$$

and

$$\frac{l_z}{l_y} \frac{V}{\bar{v}_e} \sim \frac{1}{4} \left\{ 1 - \frac{a}{l_z/2} \left[1 + \frac{kT}{e\phi_0} \right] \right\} \quad (8)$$

If $l_z/l_y = 1$ and $kT = 0.1$ ev,

$$\begin{aligned} a &\sim (1 - 4V/\bar{v}_e)(1 - kT/e\phi_0)(l_z/2) \\ a &\sim 0.85(l_z/2) \\ \phi_0 &\sim -10^{-8} HV(0.85)(l_z/2) \end{aligned} \quad (9)$$

and one end of the satellite is essentially grounded while the other end is at a potential of $-10^{-8} HV l_z$ volts, roughly 0.22 volt/m of length.

The ion energy with respect to the satellite is about 5 ev because of the satellite motion through the medium. As discussed by Jastrow and Pearse, the ion drag per incident ion will be appreciably greater than the neutral drag per incident atom for satellite potentials of several tens of volts. Moreover, the ion current will be slightly greater than the left-hand side of equation 2, as one end of the satellite drops below a potential of -5 ev. As Jastrow and Pearse show, though, even for a potential of -30 volts (i.e., appropriate to a satellite length given above of

140 meters), the increased ion collision distance of the satellite is only 9 cm, which results 0.1 per cent increase in ion current in the problem treated here. For the large satellite, where equation 8 is valid, the result that the potential end is at nearly ground potential and the average potential is proportional to the size of the satellite is especially easy to understand. Whereas the ion current is uniform over the entire projected satellite area on the plane perpendicular to the motion, the equal and opposite electron current is on a small area at the nearly-ground-potential end of the satellite. The electron flux per unit area at the ground-potential end is linearly proportional to the average electron velocity; the ion flux per unit area over the entire length is proportional to the satellite velocity. For equal total fluxes, then, it would be expected that the size of the area near or above ground potential which is responsible for the electron current would be proportional to the frontal area of the satellite times the ratio of satellite velocity to the average electron velocity. Since this velocity ratio and therefore this area is very small, the potential of the satellite midpoint will be proportional to half the length of the satellite—times the distance of the midpoint from the area at ground potential.

IV. Magnetic drag and auxiliary power possibilities. To determine the resulting satellite drag, instead of the integral of the current treated in the foregoing, the integral of ϕ times the current over the surface of the satellite must be taken. The ion current being uniform over the entire satellite, the added drag due to the satellite potential averages out to the total ion current times the average potential, ϕ_0 . The drag resulting from this constant potential has already been discussed by Jastrow and Pearse, who note that, if the impacting ions are neutralized and are then reflected, the resulting momentum transfer per ion may be significantly greater than the momentum transfer per atom for the neutral drag. In the absence of any further understanding of the accommodation coefficient at this time, the momentum of the reflected ions cannot be known. However, the ions may be neutralized and bounce off the surface with their initial 5 volts energy plus the energy acquired in dropping through the satellite potential at that point. If significant, this effect would produce a torque that would contribute

tumbling unless the neutralized particle were deflected isotropically.

We will consider three separate cases, for satellites of different lengths:

(1) $l_x/2 < a$:

$$P = ne l_y \sqrt{\frac{2kT}{\pi m_e}} \int_{-l_x/2}^{l_x/2} \left\{ \phi_0 + 10^{-8} H V x \right\} \cdot \exp \left\{ \frac{e\phi_0}{kT} \right\} \exp \left\{ \frac{e 10^{-8} H V x}{kT} \right\} dx \quad (10)$$

The first term in the brace, ϕ_0 , only compensates for the positive ion current; whatever work is done on the ions is exactly compensated by the work done by the electrons in falling through ϕ_0 . This term will be ignored in what follows.

$$\begin{aligned} P &= -n l_y \sqrt{\frac{2kT}{\pi m_e}} e^{e\phi_0/kT} \cdot \left\{ \frac{2(kT)^2}{e 10^{-8} H V} \sinh \frac{e 10^{-8} H V l_x}{2kT} \right. \\ &\quad \left. - k T l_x \cosh \frac{e 10^{-8} H V l_x}{2kT} \right\} \\ &= -2n l_y \sqrt{\frac{2kT}{\pi m_e}} k T e^{e\phi_0/kT} \cdot \left\{ \frac{kT}{e\phi_0} a \sinh \frac{e\phi_0}{kT} \frac{l_x/2}{a} \right. \\ &\quad \left. - \frac{l_x}{2} \cosh \frac{e\phi_0}{kT} \frac{l_x/2}{a} \right\} \quad (11) \end{aligned}$$

(2) $l_x/2 \ll a$:

$$P \sim \frac{2}{3} n l_y \sqrt{\frac{2kT}{\pi m_e}} \frac{(e\phi_0)^2}{kT} e^{e\phi_0/kT} \frac{(l_x/2)^3}{a^2}. \quad (12)$$

(3) $l_x/2 > a$:

$$\begin{aligned} P &= n l_y \sqrt{\frac{2kT}{\pi m_e}} \left\{ \frac{-e\phi_0}{2a} \left[\left(\frac{l_x}{2} \right)^2 - a^2 \right] \right. \\ &\quad \left. + k T a \left[1 + \frac{kT}{e\phi_0} \right] + k T \right. \\ &\quad \left. \cdot \exp \left\{ \left[\frac{e\phi_0}{kT} \left(1 + \frac{l_x/2}{a} \right) \right] \right\} \left\{ \left[\frac{l_x}{2} - \frac{kT}{e\phi_0} a \right] \right\} \right\} \\ &\sim n l_y \sqrt{\frac{2kT}{\pi m_e}} (-e\phi_0) \left\{ \frac{1}{2a} \left[\left(\frac{l_x}{2} \right)^2 - a^2 \right] \right. \\ &\quad \left. - \frac{kT}{e\phi_0} a \right\} \sim -e\phi_0 n l_y V l_x \quad (13) \end{aligned}$$

for $l_y = l_z$ (i.e., the total current, $en l_y l_z V$, times the potential drop for the electrons compared with the average satellite potential). The first term in the brace is essentially $(l_x/2 - a)$ and represents the power loss resulting from the electron current to the narrow part of the satellite which is at positive potential. The second term is the $I\Phi$ loss to all the remainder of the surface. Note that since $e\phi_0$ is in this case proportional to l_x (see equation 9) the power loss is proportional to the cube of the satellite dimension in contrast to the ordinary drag dissipation which is proportional to the square of the satellite dimension.

The ordinary drag dissipation is $\frac{1}{2} m_i \cdot V^2 N_n V l_z l_x$, which for very large satellites is $N_n \frac{1}{2} m_i V^2 / n_e e \phi_0$ of the inductive power dissipation, where N_n is the atom plus ion concentration and n_e is the electron concentration. This ratio becomes less than 1 for satellites as large as 50 meters above 1200-km altitude, where the number density of ions exceeds that of neutral particles. The source of the momentum transfer or retarding force on the satellite due to neutral or ion drag is obvious. The source of the inductive drag is the magnetic force on the satellite exerted by the external magnetic field due to the electric current within the satellite metal.

$$\mathbf{F} = \int_{-l_x/2}^{l_x/2} 10^{-8} I(x) \, d\mathbf{x} \times \mathbf{H} \quad (14)$$

Since

$$\begin{aligned} I(x) &= en l_y \int_{-l_x/2}^x \left\{ V - \sqrt{\frac{2kT}{\pi m_e}} \right. \\ &\quad \left. \cdot \exp \left[\frac{e\phi_0}{kT} \left(1 - \frac{x'}{a} \right) \right] \right\} dx' \quad \text{for } x < a \quad (15) \end{aligned}$$

and

$$\begin{aligned} I(x) &= en l_y \left(V - \sqrt{\frac{2kT}{\pi m_e}} \right) \\ &\quad \cdot \int_x^{l_x/2} dx' \quad \text{for } x > a \end{aligned}$$

for $l_x/2 > a$,

$$\begin{aligned} F &= \int_{-l_x/2}^{l_x/2} I(x) 10^{-8} H \, dx \\ &= 10^{-8} H en l_y \left\{ \sqrt{\frac{2kT}{\pi m_e}} \frac{1}{2} (l_x/2 - a)^2 \right. \end{aligned}$$

$$\begin{aligned}
& + V a l_x + \sqrt{\frac{2kT}{\pi m_e}} e^{e\phi/kT} a \left(\frac{kT}{-e\phi_0} \right) \\
& \cdot \left[\exp \left\{ \frac{e\phi_0}{kT} \left(\frac{l_x/2}{a} \right) \right\} \left(\frac{l_x}{2} + a - a \frac{kT}{e\phi_0} \right) \right. \\
& \left. + a \frac{kT}{e\phi_0} e^{e\phi_0/kT} \right] \} \quad (16) \\
& \sim e\phi_0 n l_y l_x \quad (13a)
\end{aligned}$$

Of course (16) is identical to (13), as a tedious arithmetical exercise can show. Thus, if the neutral density is less than the ion density, as it is above 1200-km altitude, and the satellite is larger than roughly 50 meters, the inductive drag will exceed the ordinary neutral or ion drag. This is quite reasonable in that the electrons in dropping through a potential greater than 5 volts more than the average ions do are losing more energy than the impacting ions gain. This is of no consequence for currently contemplated satellites, which are either too small or, if this large, too heavy to experience appreciable drag at altitudes in excess of 1200 km. Only for large conducting 'balloons' will the inductive drag be a dominant or even appreciable effect.

A satellite of the order of 100 meters in diameter could generate perhaps 100 watts of power from inductive effects if its skin were composed of electrically conducting strips separated by insulating layers, so that the induced current would be caused to flow through an electric motor or whatever. The useful power figure of 100 watts is determined by the ion current to the satellite which limits the total inductive power to 200 watts at a maximum ion density of $10^6/\text{cc}$ at 400-km altitude. If electrons could ever be caused to be emitted from the negative end of the satellite, this power could be increased by an order of magnitude, but it can never be an important power source in terms of the auxiliary power likely to be required by an ordinary satellite large enough to generate significant power. Since the neutral drag at this altitude far exceeds the inductive drag, useful generation of electrical power from the satellite motion will not appreciably shorten the satellite lifetime.

V. Drift currents, photoelectrons, and antennas.

In the preceding sections only the motion of electrons parallel to the magnetic field has been considered. Motion perpendicular to the mag-

netic field is very much reduced in the absence of collisions, but in an applied electric field perpendicular motion does give rise to tin drift currents. The force on a charge in electric and magnetic fields is given by

$$m \frac{d\mathbf{v}}{dt} = e(\mathbf{E} + \mathbf{v} \times \mathbf{H}) \quad (17)$$

Letting $\mathbf{v} = \mathbf{v}' + \mathbf{E} \times \mathbf{H}/H^2$, where \mathbf{E} and \mathbf{H} are constant in time,

$$m \frac{\partial \mathbf{E}}{\partial S} \frac{dS}{dt} \frac{\mathbf{E} \times \mathbf{H}}{EH^2} + m \frac{d\mathbf{v}'}{dt} = e(\mathbf{v}' \times \mathbf{H}) \quad (18)$$

The first term and the $\mathbf{E} \times \mathbf{H}/H^2$ part of \mathbf{v} give rise to any drift currents. If the satellite were charged, a drift current would encircle the satellite, causing a small magnetic dipole parallel to or opposed to the external field, depending on the charge of the satellite. As for the induced field perpendicular to both the satellite motion and the magnetic field, it causes drift currents parallel to and antiparallel to the satellite velocity. But all these effects can be neglected in comparison with the inductive drag discussed previously. Intuitively, one would anticipate this result, knowing that plasma conductivity or charge mobilities are several orders of magnitude higher along the magnetic field than perpendicular to it. One negligible modification should be made in our previous work, in that the effective centers of gyration of the electrons and ions are slightly shifted toward lower electric potential which has the effect of very slightly altering the effective length of the satellite in the direction perpendicular to both the satellite motion and the magnetic field.

Photoelectrons and secondary electrons from conceivable satellite surfaces are too few to modify any of the results presented here.

The orientation of antennas will make no difference with respect to antenna potential and the negligible inductive drag. If an antenna were perpendicular to both satellite motion and magnetic field, the satellite dimension in that direction would be effectively increased. The antennas may not change the total current appreciably, but they do increase the potential through which the discharge current falls, and the inductive drag is quadratically rather than linearly dependent on this dimension of the satellite.

VI. *Conclusions.* In the absence of satellite motion perpendicular to a magnetic field, a satellite accumulates a positive charge on its front surface by running into slowly moving positive ions, and a negative charge from being bombarded by fast-moving electrons. In the absence of a magnetic field, the electrons are incident on all satellite surfaces, and they are incident on those surfaces perpendicular to the field if a field exists. Since the electron velocity v exceeds the satellite velocity, a higher incident flux of negative charges will result unless a sufficiently large negative satellite potential is formed to repel the excess negative electron current. The resulting negative satellite potential, given exactly by equation 5, is approximately $6 kT/e$, which amounts to 0.25 volt.

For satellite motion perpendicular to a magnetic field, the above result is modified by an induced emf. For satellites larger than about 10 meters (depending on the strength of the magnetic field and the direction of satellite motion with respect to the field) one end becomes essentially grounded, owing to a high incident flux of negative electrons there, while the other end is at a negative potential of about 0.2 volt per meter of satellite dimension perpendicular to the magnetic field. The positive ion current is

uniform over the frontal area of the satellite, but the negative electron current is all at the ground-potential end. The resulting electric current within the satellite is acted on by the external magnetic field, causing a retarding force which is equal to the electric power lost by the incident charges divided by the satellite velocity. As this inductive drag is very much less than the neutral drag at all altitudes below 1200 km and is insignificant above that altitude, it is generally not an appreciable effect. However, very small amounts of induced electrical power of roughly a few hundred watts for surfaces 100 meters in diameter are available as an auxiliary power source at altitudes of around 200 to 800 km.

The authors are grateful to Drs. A. J. Dessler and W. B. Hanson of Lockheed for illuminating conversations and an informed and helpful reading of the manuscript.

REFERENCE

- Jastrow, R., and C. A. Pearce, Atmospheric drag on the satellite, *J. Geophys. Research*, 62, 413-423, 1957.

(Manuscript received August 3, 1959;
revised October 15, 1959.)

Satellite Orbits in an Oblate Atmosphere

D. G. PARKYN

*Department of Mathematics
University of Cape Town, South Africa*

Abstract. Density values of the upper atmosphere derived from the rate of development of the orbits of earth satellites suffer from errors introduced by uncertainties in the determination of the mean drag cross section. As a result it is difficult or impossible to detect minor variations in the density such as, for example, its latitude dependence. An analysis is developed which circumvents this difficulty by considering the behavior of a dynamically simple parameter—the square of the product of the semimajor axis and the eccentricity. Examples from published orbital data are discussed, showing clearly the presence of a latitude dependence, and estimates of the ratio of polar to equatorial density are deduced.

INTRODUCTION

There is a fairly extensive literature on the effect of atmospheric drag on the orbit of an earth satellite and on the consequent deduction of atmospheric conditions from observational data. The current methods follow two essentially distinct schools. The first derives density values from the rate of change of periodic time, the orbital acceleration [Schilling and Whitney, 1958]; the second analyzes the long-term effects of a given atmospheric model and by comparison deduces atmospheric properties [Parkyn, 1958a and b; King-Hele and Leslie, 1958]. The first method suffers from the defect that the ultimate possible error in the derived densities is large enough to preclude any determination of minor density variations resulting, for example, from the oblateness of the earth. The second method, on the other hand, leads to results that are independent of the drag cross section, apart from the requirement that it be effectively constant, and hence removes one of the major possible errors in the density determination. It is proposed to extend the analysis for a spherical atmosphere of King-Hele and Leslie [1958] and Parkyn [1958b] to the case of an oblate atmosphere and to investigate the dynamical effects of such a model. It will be found that latitude variations in the density at high levels are readily observable, and estimates of these variations will be made from orbital data.

EXTENSION OF THE SPHERICAL ANALYSIS

Since the effects to be considered are of the order e_0^2 , where e_0 is the initial eccentricity of the

orbit, it is necessary to extend the existing analysis for the spherical case. Adopting the standard results for small perturbations [Parkyn, 1958b], we find that the change in semimajor axis (Δa) and the change in the eccentricity (Δe) are given by

$$\Delta a = -4\pi a^2 \{ (1 + e^2)F + 2eG \}$$

$$a \Delta e = -4\pi a^2 (1 - e^2) \{ G + eF \}$$

where F and G represent the dependence of the variations Δa and Δe upon the spread of the drag from the perigee point. More explicitly,

$$F = \frac{1}{2\pi} \int_{-\pi}^{\pi} k(r) \cdot \frac{(1 + 2e \cos \psi + e^2)^{1/2}}{(1 + e \cos \psi)^2} d\psi$$

$$G = \frac{1}{2\pi} \int_{-\pi}^{\pi} k(r) \cdot \frac{(1 + 2e \cos \psi + e^2)^{1/2}}{(1 + e \cos \psi)^2} \cos \psi d\psi$$

where r is the radial distance of the satellite from the center of the earth and ψ the angular distance of the satellite from perigee. The term $k(r)$ is a drag factor dependent upon the density $\rho(r)$, the effective cross section S , the mass M , and the aerodynamic drag coefficient C_D of the satellite, such that

$$k(r) = \frac{1}{2}(C_D S / M) \rho(r) \quad (1)$$

If we assume that the atmosphere may be approximated isothermally with a scale height H in the neighborhood of the perigee point, then

$$F = \frac{k(p)}{2\pi} \int_{-\pi}^{\pi} \exp\left(-\frac{r-p}{H}\right) \cdot \frac{(1+2e \cos \psi + e^2)^{1/2}}{(1+e \cos \psi)^2} d\psi$$

$$G = \frac{k(p)}{2\pi} \int_{-\pi}^{\pi} \exp\left(-\frac{r-p}{H}\right) \cdot \frac{(1+2e \cos \psi + e^2)^{1/2}}{(1+e \cos \psi)^2} \cos \psi d\psi$$

where p is the radial distance of the perigee from the center of the earth.

For small values of the eccentricity it is convenient to expand the equation of the orbit in the form

$$r = p \left\{ 1 + \left(\frac{e}{1-e} \right) (1 - \cos \psi) + \left(\frac{e}{1+e} \right)^2 (1 - \cos \psi)^2 + \dots \right\}$$

If we write $\mu = pe/(1+e)$, then, correct to the first power of e ,

$$F = \frac{1}{\pi} \exp\left(-\frac{\mu}{H}\right) \int_0^{\pi} \exp\left(\frac{\mu \cos \psi}{H}\right) \cdot (1 - e \cos \psi) d\psi$$

$$G = \frac{1}{\pi} \exp\left(-\frac{\mu}{H}\right) \int_0^{\pi} \exp\left(\frac{\mu \cos \psi}{H}\right) \cdot (1 - e \cos \psi) \cos \psi d\psi$$

Hence,

$$F = \exp\left(-\frac{\mu}{H}\right) \left\{ I_0\left(\frac{\mu}{H}\right) - e I_1\left(\frac{\mu}{H}\right) \right\}$$

$$G = \exp\left(-\frac{\mu}{H}\right) \left\{ \left(1 - \frac{H}{p}\right) I_1\left(\frac{\mu}{H}\right) - e I_2\left(\frac{\mu}{H}\right) \right\}$$

where I_0, I_1 are the zero and first-order Bessel functions of purely imaginary argument.

It is convenient to consider the variations of A ($= ae$) and p rather than a and e where

$$\begin{aligned} \Delta A &= \Delta(ae) \\ &= -4\pi a^2 \{G + 2eF\} \end{aligned}$$

and

$$\Delta p = -4\pi a^2 \{F - G\} (1 - 2e)$$

to a sufficient order. Further, taking a reference level p_0 near the level of the perigee, for example at the initial level of p ,

$$k(p) = k(p_0) \exp\left(-\frac{p-p_0}{H}\right)$$

on the isothermal assumption.

Defining the mean rates of change of A and p as $\Delta A/T$ and $\Delta p/T$ respectively, where T is the anomalistic period, we find on substitution

$$\begin{aligned} \dot{A} &= K^* a^{1/2} \exp\left(-\frac{p+\mu}{H}\right) \cdot \left\{ \left(1 - \frac{H}{p}\right) I_1\left(\frac{\mu}{H}\right) + e I_0\left(\frac{\mu}{H}\right) \right\} \quad (2) \end{aligned}$$

$$\begin{aligned} \dot{p} &= K^* a^{1/2} \exp\left(-\frac{p+\mu}{H}\right) \cdot \left\{ \left(I_1\left(\frac{\mu}{H}\right) - I_2\left(\frac{\mu}{H}\right) \right) (1 - e) + \frac{H}{p} I_1\left(\frac{\mu}{H}\right) \right\} \quad (3) \end{aligned}$$

where

$$K^* = -\pi \frac{C_D S}{M} \sqrt{g R^2} \exp\left(\frac{p_0}{H}\right) \rho(p_0)$$

R being the mean radius of the earth. K^* will be assumed to remain constant, a restriction implying that the mode of tumbling remains unchanged and that the period of such motion is small compared with the periodic time.

SOLUTION OF THE EQUATIONS

It is possible to solve equations 2 and 3 for the early part of an orbit by expanding the Bessel functions in their asymptotic expansions [Leslie, 1958], but such solutions become invalid relatively early in the orbital development. As an alternative procedure we make use of the derivative properties of the Bessel functions to form alternative differential equations. Thus

$$\frac{\dot{p}}{\dot{A}} = \frac{\left[I_1\left(\frac{\mu}{H}\right) - I_2\left(\frac{\mu}{H}\right) \right] (1 - e) + \frac{H}{p} I_1\left(\frac{\mu}{H}\right)}{\left(1 - \frac{H}{p}\right) I_1\left(\frac{\mu}{H}\right) + e I_0\left(\frac{\mu}{H}\right)}$$

and on differentiating and using relations 2 and 3 to simplify we find that

$$\begin{aligned} \frac{d}{dt} \left\{ \frac{\dot{p}}{\dot{A}} \right\} &= \frac{\dot{\mu}}{H} \left\{ \frac{H}{\mu} \left(1 - \frac{2H}{p} \right) \right. \\ &\quad - \left(\frac{\dot{p}}{\dot{A}} \right) \left(2 - \frac{H}{\mu} + \frac{H^2}{\mu^2} - \frac{H^2}{\mu p} \right) \\ &\quad \left. - \left(\frac{\dot{p}}{\dot{A}} \right) \left(1 + \frac{\mu}{p} \right) \right\} \quad (4) \end{aligned}$$

Hence, writing $\mu/H = m$, we see that (d/dm) (\dot{p}/\dot{A}) is a function of m and \dot{p}/\dot{A} with the exception of two terms. The first, $H^2/\mu p$, can be adequately approximated by writing p_0 for p ; the second, $(\dot{p}/\dot{A})^2(\dot{\mu}/p)$, is sufficiently small throughout the life span to be safely ignored. Thus (\dot{p}/\dot{A}) can be adequately treated as a function of m , which we will write as $f(m)$, and equation 4 can be written

$$\frac{df}{dm} = \frac{1}{m} \left(1 - \frac{2H}{p_0} \right) - f \left[2 - \frac{1}{m} \left(1 + \frac{H}{p_0} \right) + \frac{1}{m^2} \right] - f^2 \quad (5)$$

This is a Riccati equation and can be solved in series for small or large m . In practice m ranges from 10 or more to 0, but the range 1 to 0 occupies only the last day or so of the life span. Hence it is effective to adopt an expansion in inverse powers of m for the greater part of the satellite's life. Thus

$$f \approx \frac{1}{2m} + \frac{3}{8m^2} + \frac{1}{8m^3} + \dots \quad (6)$$

Giving as a first approximation

$$p - p_0 = \frac{H}{2} \ln \frac{A}{A_0}$$

where p_0 is the initial value of p . For a second approximation

$$\begin{aligned} \frac{H}{\mu} &= \frac{H(1+e)}{pe} = \frac{H}{A} \left(\frac{1+e}{1-e} \right) \\ &\approx \frac{H}{A} \left(1 + \frac{2A}{p_0} \right) \end{aligned}$$

Hence

$$\begin{aligned} \frac{p - p_0}{H} &\approx \frac{1}{2} \ln \frac{A}{A_0} + \frac{A - A_0}{p_0} \\ &\quad - \frac{3H}{8} \left(\frac{1}{A} - \frac{1}{A_0} \right) \quad (7) \end{aligned}$$

affording a direct method of determining the scale height at the level of the perigee which is independent of the drag cross section [cf. King-Hele, 1959].

The second alternative differential equation can be formed by noting that

$$\mu \ddot{A} + \dot{\mu} \dot{A} \approx -\frac{5}{2} \frac{\dot{A} \dot{\mu}}{p_0} - \frac{3}{4} \frac{H \dot{A} \dot{\mu}}{p_0} \quad (8)$$

correct to terms in $1/p_0$. In the reduction p has been replaced by the nearly equivalent p_0 and use has been made of the approximate relation 6. Since $\mu \sim A$ and the terms on the right-hand side of (8) are small we have, as a first approximation,

$$A \ddot{A} + (\dot{A})^2 \approx 0 \quad (9)$$

showing that A^2 is very nearly a linear function of time. This property is the basis of estimates of a satellite's life [Scott, 1957; Leslie, 1958].

However, this approximation fails at points near the end of the life span, roughly when t/t_{final} exceeds 0.9, when it is found that the graph of A falls below the linear form. It is possible to extend the range of validity of the solution of (9) by integrating equation 8, although such an extension cannot have complete validity since it depends on the approximate relation 6.

We find that

$$A^{1+\tau} \dot{A} \approx \frac{2A^{2+\tau} \dot{A}}{p_0} + C \left(1 - \frac{5}{2} \frac{A}{p_0} \right) \quad (10)$$

on approximating to μ in terms of A and p_0 and writing $\tau = 3H/4p_0$. Further, if we replace the variable A by $v = A/A_0$, where A_0 is the initial value of A , and make use of the approximation that

$$v^2 = 1 - Kt$$

in the small terms on the right of (10) we find that

$$\begin{aligned} v^2 &\approx [(1 - Kt) \\ &\quad + \frac{1}{3} e_0 \{ 1 - (1 - Kt)^{3/2} \}]^{1-\tau/2} \quad (11) \end{aligned}$$

Since τ is small, v^2 is effectively a linear function of t over most of the life, but the extended solution has the advantage of fitting the final stages very much more closely. The obvious advantage of the parameter v^2 over the observationally more direct periodic time is that its near-linear behavior makes any anomalies easier to determine.

THE EFFECT OF AN OBLATE ATMOSPHERE

The problem of orbital development in an oblate atmosphere will be treated by a perturbation technique, using the solution (11) as a first approximation to the behavior of v^2 . We

commence by determining the density distribution in an isothermal rotating atmosphere.

The hydrostatic equilibrium condition leads to the relation

$$\gamma \ln \rho = \frac{gR^2}{r} + \frac{\nu}{r^3} (\cos^2 \theta - \frac{1}{3}) + \frac{1}{2} \omega^2 r^2 \sin^2 \theta \quad (12)$$

for the density ρ at a point whose distance from the center of the earth is r and whose angular distance from its axis is θ . In this relation R is the mean radius of the earth, ω its angular velocity, and γ , dependent upon the molecular weight of the atmosphere, will be assumed constant. Further,

$$\frac{\nu}{R^3} = \frac{1}{2} \omega^2 R^2 - \epsilon g R$$

where ϵ is the ellipticity of the earth.

If we write $R + z = r$ and assume that $z \ll R$, we find on reduction that

$$\rho = \rho_0 \exp \left\{ -\frac{z}{H} \left[1 - \frac{1}{2} \epsilon - \frac{1}{2} \epsilon \cos^2 \theta \right] - \frac{\epsilon R \cos^2 \theta}{H} \right\} \quad (13)$$

where $H = \gamma/g$ and the approximate relation $\epsilon g R \approx \omega^2 R^2$ has been used in the reduction. This would imply for the θ dependence of the density at the reference level p_0

$$\rho(p_0, \theta) = \rho_0' \cdot \exp \left\{ -\left(\frac{3R - p_0}{2H} \right) \epsilon \cos^2 \theta \right\} \quad (14)$$

We shall assume that the density at any level has a θ dependence in this form, but shall avoid any immediate assumption of values for the constants, writing

$$\rho(p_0, \theta) = \rho(p_0) \exp(-\lambda \cos^2 \theta) \quad (15)$$

In the neighborhood of the level p_0 we shall assume that the density varies as in equation 13, simplifying it to the form

$$\rho(z, \theta) = \rho(p_0, \theta_0) \cdot \exp \{ -(z/H) - \kappa \cos^2 \theta \} \quad (16)$$

where z is measured from the level of the perigee, whose coordinates are p_0, θ_0 . Hence for two

neighboring points $(z, \theta), (z', \theta')$ we have

$$\frac{\rho(z', \theta')}{\rho(z, \theta)} = \exp \left\{ -\frac{z' - z}{H} - \kappa (\cos^2 \theta' - \cos^2 \theta) \right\} \quad (17)$$

Finally, using the proportionality of the drag factor k to the density, we have in an obvious notation

$$\begin{aligned} k(r, \theta) &= k(p_0, \theta_0) \cdot \frac{k(p, \theta_0)}{k(p_0, \theta_0)} \cdot \frac{k(r, \theta)}{k(p, \theta_0)} \\ &= k_0 \exp[-\lambda \cos^2 \theta_0] \cdot \exp \left[-\frac{p - p_0}{H} \right] \\ &\quad \times \exp \left[-\frac{r - p}{H} - \frac{\kappa}{2} (\cos 2\theta - \cos 2\theta_0) \right] \end{aligned}$$

where k_0 is a constant dependent purely upon the choice of reference level p_0 .

Having assumed an arbitrary but ostensibly suitable density distribution in the neighborhood of perigee we may proceed with the evaluation of the changes in the orbital elements in one revolution. We shall restrict the discussion to the change in A ($= ae$) and assume that the perigee distance behaves as for a spherical atmosphere.

The change in A in one revolution of the satellite in orbit is given by

$$\Delta A = -4\pi a^2 [G' + 2eF']$$

where

$$F' = \frac{1}{2\pi} \int_{-\pi}^{\pi} k(r, \theta) \cdot \frac{(1 + 2e \cos \psi + e^2)^{1/2}}{(1 + e \cos \psi)^2} d\psi$$

and

$$G' = \frac{1}{2\pi} \int_{-\pi}^{\pi} k(r, \theta) \cdot \frac{(1 + 2e \cos \psi + e^2)^{1/2}}{(1 + e \cos \psi)^2} \cos \psi d\psi$$

the angle ψ again being measured from perigee. Further, if α is the inclination of the orbital plane to the equator and β is the angle from apogee to perigee, measured positively in the direction of rotation of the earth,

$$\cos \theta = \sin \alpha \cos (\psi + \beta)$$

The integrals F' , G' may be evaluated by expanding the integrands in powers of the small parameter κ . In fact, it is sufficient to retain only the first power of κ for a first-order oblateness correction. After reduction we find that

$$F' \sim k_0 \exp\left(\frac{p_0}{H}\right) \exp\left(-\frac{p+\mu}{H}\right) \\ \times \left[I_0\left(\frac{\mu}{H}\right) \cdot (1 - \lambda \sin^2 \alpha \cos^2 \beta) \right. \\ \left. - I_1\left(\frac{\mu}{H}\right) \cdot \left(e - \frac{\kappa H}{\mu} \sin^2 \alpha \cos 2\beta\right) \right]$$

and

$$G' \sim k_0 \exp\left(\frac{p_0}{H}\right) \exp\left(-\frac{p+\mu}{H}\right) \\ \times \left[I_1\left(\frac{\mu}{H}\right) \left(1 - \frac{H}{p} - \frac{2\kappa H^2}{\mu^2} \sin^2 \alpha \cos 2\beta\right) \right. \\ \left. + I_0\left(\frac{\mu}{H}\right) \cdot \left(-e + \frac{\kappa H}{\mu} \sin^2 \alpha \cos 2\beta\right) \right]$$

If we denote the solution of the corresponding spherical problem by the subscript unity, we may replace the Bessel functions by A_1 and its derivative.

Thus, to a sufficient order,

$$\dot{A} = \dot{A}_1 \left(1 - \frac{\lambda}{2} \sin^2 \alpha\right) \\ - \frac{\lambda \dot{A}_1}{2} \sin^2 \alpha \cos 2\beta \\ + \kappa \sin^2 \alpha \left[\frac{H \dot{A}_1}{A_1} \cos 2\beta \right. \\ \left. - \frac{3}{2} \frac{H^2}{A_1^2} \dot{A}_1 \cos 2\beta \right] \quad (18)$$

To integrate this relation we note that β increases approximately linearly with time, and so we shall write $\beta = \omega t + \beta_0$, where ω is the mean rate of increase of β . Further, in the correction terms it is sufficient to substitute the approximation

$$A_1 = A_0(1 - Kt)^{1/2}$$

where A_0 again refers to the initial value of A . It is convenient to define

$$\beta_1 = \beta_0 + (\omega/K)$$

where, since $1/K$ is an approximation to the life span, β_1 may be interpreted as the final value of β . On integrating equation 18 we find that

$$v = A/A_0 \\ = \left\{ v_1 \left(1 - \frac{\lambda}{2} \sin^2 \alpha\right) + \frac{\lambda}{2} \sin^2 \alpha \right\} \\ - \frac{\lambda}{4} \sqrt{\frac{\pi K}{|\omega|}} \{ \cos 2\beta_1 [C(2|\beta - \beta_1|) \\ - C(2|\beta_0 - \beta_1|)] \\ \pm \sin 2\beta_1 [S(2|\beta - \beta_1|) \\ - S(2|\beta_0 - \beta_1|)] \} \\ + \frac{\kappa H}{2A_0} \sin^2 \alpha \{ \cos 2\beta_1 [Ci(2|\beta - \beta_1|) \\ - Ci(2|\beta_0 - \beta_1|)] \\ \pm \sin 2\beta_1 [Si(2|\beta - \beta_1|) \\ - Si(2|\beta_0 - \beta_1|)] \} \quad (19)$$

The functions S , C and Si , Ci are the Fresnel and exponential sine and cosine integrals, respectively. The upper or lower signs are to be taken as ω is positive or negative.

A SIMPLIFIED FORM FOR A YOUNG ORBIT

When $\beta - \beta_1$ is large, that is for small values of t , the integral functions may be replaced by their asymptotic expansions to produce a simpler form of (19). This result can be accomplished more directly, however, by evaluating the original integrals by continued integration by parts. The result is effectively that

$$v \sim v_1 + \left(\frac{K}{4\omega}\right) \cdot \left(\frac{\lambda}{2} - \frac{\kappa H}{A_0}\right) \sin^2 \alpha \sin 2\beta \quad (20)$$

or, since t is small and v_1 near unity,

$$v^2 \sim v_1^2 + \left(\frac{K}{2\omega}\right) \left(\frac{\lambda}{2} - \frac{\kappa H}{A_0}\right) \sin^2 \alpha \sin 2\beta \quad (21)$$

For a completely isothermal atmosphere $\lambda = (3R - p_0)\epsilon/2H$ and $\kappa = R\epsilon/H$, so that we may expect λ and κ to be comparable. Therefore for an orbit of moderate eccentricity (~ 0.1) the term in κ leads to a correction of the order of 10 per cent to the amplitude of the resulting oscillations. In view of the relative magnitude of the effect considered it is therefore reasonable

to ignore the term κ , equivalent to assuming that the scale height above the reference level p_0 is independent of θ . Hence in discussing, for example, tidal disturbances it is sufficient in a preliminary investigation to postulate only the density variation at the level p_0 .

APPLICATIONS OF THE ANALYSIS

Form of the periodic time curve. From the relations (7), (11), and (19) the semimajor axis may be determined and thence the periodic time as a function of t . In view of the complexity of the form of such a relation, even under the most drastic simplifications, however, it would seem desirable to study orbital data with reference to equation (19), or preferably the simplified form (21) when Kt is small.

Determinations of mean density. The form of A^2 as a function of t as determined observationally is that of small, sometimes irregular, oscillations about a curve given by (11). Indeed, for values of Kt less than about 0.3 the mean curve is closely linear. Hence the constant K can be accurately determined and affords a method of determining the mean density at perigee level.

Thus, expanding the Bessel functions in equation (2) in asymptotic series we find that

$$\dot{A}_0 \sim -2k(p_0) \sqrt{\frac{gR^2 H}{2\pi e_0}} \cdot (1 + e_0)$$

whereas, if the mean curve can be approximated by $v^2 = 1 - Kt$,

$$\dot{A}_0 = -\frac{1}{2} A_0 K$$

Hence the density at perigee, ρ_π , is given by

$$\rho_\pi = \frac{1}{2} \frac{KM}{C_D S} \left(\frac{a_0}{1 + e_0} \right) \sqrt{\frac{gR^2 H e_0}{2\pi}}$$

where the constants of the satellite are defined as in (1). This is equivalent to the relations of *Sterne* [1958] and *King-Hele* [1959] and requires, as do their results, a knowledge of the scale height H . It may be modified in the same way to make this dependence negligible by using it to determine the mean density at a height $\frac{1}{2}H$ above the perigee. As an alternative method it has the advantage of depending upon the accurately determinate K whereas the orbital acceleration is subject to large fluctuations [*Jacchia and Briggs*, 1958; *King-Hele and Walker*, 1959].

Determination of the scale height. It has been noted that the relation 7 offers a direct method for the determination of the scale height H which is independent of the effective drag cross section and the drag coefficient [*King-Hele* 1959]. It suffers, however, from the defect that $(p - p_0)$ is difficult to determine with high accuracy. It is probable that with the accumulation of data this method will ultimately give a fairly clear picture of the variation of H with height, a knowledge which is essential to any temperature estimates.

Determination of the oblateness of the upper atmosphere. If we ignore the small term in κ in equation 21 we note that the amplitude of the oblateness oscillations is given by

$$(K\lambda/4\omega) \sin^2 \alpha \quad (22)$$

where K , ω , and α are accurately known. Hence isolation of these oscillations leads to an estimate of the ratio of the polar to the equatorial density at the level of the perigee as $\exp(-\lambda)$. A knowledge of the scale height will convert this into an estimate of the oblateness of a surface of constant density through the mean perigee level. Such an estimate, while assuming an effectively constant drag cross section, is free from the uncertainties arising in the search for latitude dependence in deduced densities where the variations are less than the probable error [*Schilling and Whitney*, 1958; cf. *Veis*, 1959].

EXAMPLES FROM ORBITAL DATA

1958 α . Data, Teske, private communication. Figure 1 is a graph of the departure from linearity of v^2 as a function of time. More precisely, the ordinate of curve 1, δ , is

$$[v^2 - (1 - Kt - C)] \times 10^6,$$

where K has the value 6.006×10^{-4} /day. Curve 2, for the oblateness term $\sin 2\beta$, has been inserted for comparison. There is evidently a close correlation between the higher-frequency oscillations and the oblateness correction, and an approximate Fourier analysis for the period 26 to 11 days requires an amplitude of 340×10^{-6} for this term. Hence from equation 22 we determine λ as 0.849, giving the ratio of the polar to the equatorial density at the initial level of the perigee as 0.43 from equation 15. This final deduction is little more than a conjecture, since

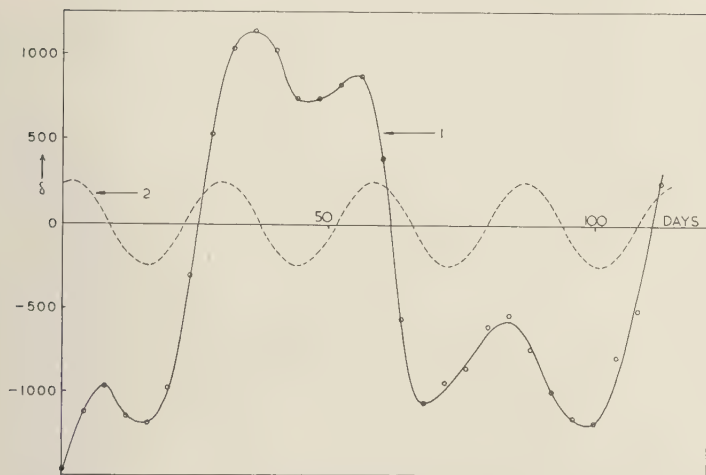


Fig. 1. Comparison of $[\nu^2 - (1 - Kt)]$ and oblateness correction for satellite 1958 α . ($t_0 = 32.16$ days after 1958.0.)

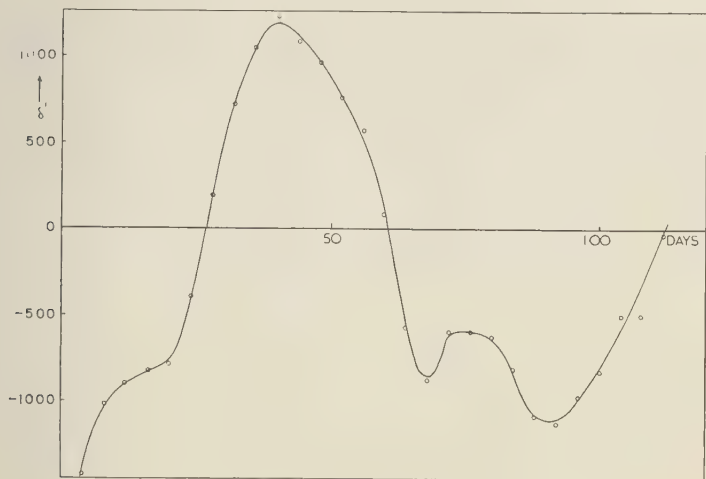


Fig. 2. Curve for 1958 α with the oblateness correction removed.

the latitude range over which λ has been determined is restricted to -32° to $+32^\circ$.

If we deduct the estimated oblateness correction from δ to determine the ordinate δ' of Figure 2 we find that the fit is by no means precise. There is evidence of a marked change in phase that cannot be explained by errors in the determination of β . Only in the section from 26 to 62 days is the correction adequate, a period during which the sun was passing over the equator. There is no reason for a purely gravitational oblateness to

show anomalies of this type, and the implication would seem to be that a significant part of the effect is a result of solar action. A full analysis must await more extensive data. The remaining wave form agrees closely with a fourth harmonic tide [Parkyn, 1959], and will be discussed in another paper.

1958 β_2 . Data, Jacchia and Briggs [1958]. Curve 1 of Figure 3 has as its ordinate

$$\delta = [\nu^2 - (1 - Kt)] \times 10^7,$$

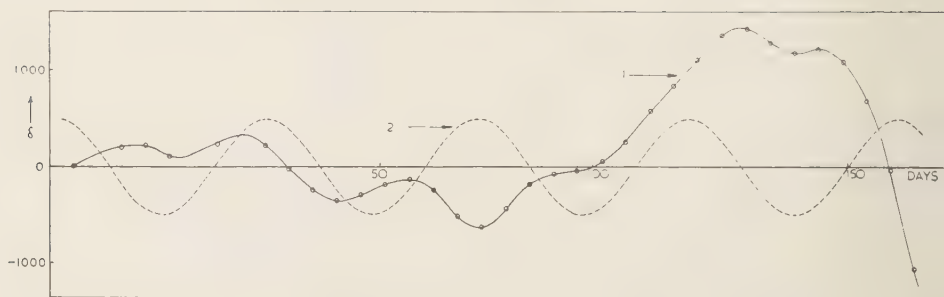


Fig. 3. Comparison of $[v^2 - (1 - Kt)]$ and oblateness correction for satellite 1958 β_2 . ($t_0 = 76.6$ days after 1958.0.)

with K as 1.527×10^{-5} /day. Curve 2 is that of $\sin 2\beta$. The oblateness effect is not particularly evident, though the form of curve 1 from 90 to 150 days suggests the presence of such an oscillation with an amplitude of the order of 150×10^{-7} . Without adequate knowledge of the form of the remainder after the removal of such a correction any accurate estimate of the amplitude would be purely conjectural. Curve 1 does, however, give a picture of the anomalies in the behavior of the orbit which is far clearer than that obtainable from the orbital accelerations [cf. *Jacchia and Briggs, 1958*].

1958 ϵ . Data, *Schilling and Whitney [1958]*. In Figure 4, curve 1 gives $\delta = [v^2 - (1 - Kt)] \times 10^5$, with K as 2.751×10^{-3} /day. Curve 2 is again of $\sin 2\beta$, and the oblateness effect is strongly marked. In Figure 5, curve 1 is that of

$$\delta' = \delta - 350 \times 10^{-5} (\sin 2\beta - \sin 2\beta_0)$$

and curve 2 is of the cosine of 3 times the difference in right ascension of the sun and the

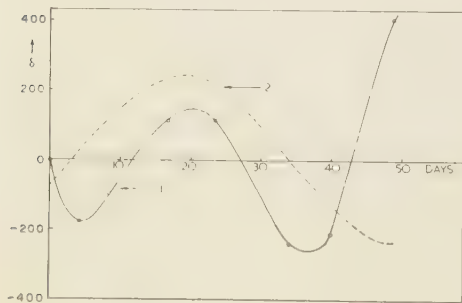


Fig. 4. Comparison of $[v^2 - (1 - Kt)]$ and oblateness correction for satellite 1958 ϵ . ($t_0 = 219.02$ days after 1958.0.)

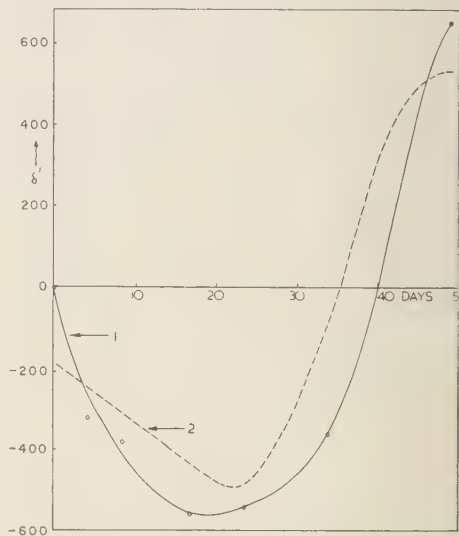


Fig. 5. Curve for 1958 ϵ with the oblateness correction removed, with curve of the cosine of three times the difference in right ascension of the satellite and the sun for comparison.

perigee point. The correlation between the two curves suggests the presence of a third-harmonic tide [cf. *Parkyn, 1959*], though the data are inadequate for a detailed analysis.

If the amplitude of the oblateness correction is taken to be 350×10^{-5} , we determine λ as 0.435 and the ratio of polar to equatorial density at a height of 160 miles to be 0.65.

CONCLUSION

The purpose of this paper has been to present an analysis of the dynamical effects of an oblateness

atmosphere on the orbit of a satellite and to show that the resulting effects are readily detectable in the available data. The examples considered have not been given with the aim of deriving accurate estimates of atmospheric oblateness at the levels concerned, since in each case better estimates will be possible with the publication of more extended orbital data. At the same time, the examples given show well the increased clarity that results from the use of a v^2 analysis of the data as compared with the orthodox discussions of the orbital acceleration.

Although none of the Russian satellites has been considered, an oblateness effect is again evident in the data, but because of the near-critical inclination of their orbital planes the resulting oscillations have a much longer period. In the case of Sputnik II (1957 β) only half an oscillation was completed in its lifetime and the effect was inextricably mixed with the intricacies of the latter stages of its orbital development. The orbit of 1958 δ_2 , however, should show the effect clearly when more data become available, since its longer life will permit several complete oscillations.

The analysis given discusses only the effect of the oblateness of the earth. It is evident that other factors may disturb the quasi-linearity of the v^2 curve. For example, there seems to be evidence of atmospheric tides [Parkyn, 1959], and again variations of perigee height due to a third harmonic in the earth's gravitational field would produce variations of the same order [O'Keefe and others, 1959]. But these are small

corrections to a basic spherical behavior and may be superposed with sufficient accuracy. Further, it must be emphasized that there are anomalies of two different orders in satellite behavior, and the method used smoothes out the more rapid fluctuations of the type noted by King-Hele [1958] in the large day-to-day fluctuations in the orbital acceleration. On the other hand, the simpler picture presented in this manner may yield ultimately a complete analysis of the long-term effects.

REFERENCES

- Jacchia, L. G., and R. E. Briggs, *Smithsonian Astrophys. Observatory Spec. Rept.* 18, 9, 1958.
 King-Hele, D. G., *Nature*, 183, 1224, 1959.
 King-Hele, D. G., and D. C. M. Leslie, *Nature*, 181, 1761, 1958.
 King-Hele, D. G., and D. M. C. Walker, *Nature*, 183, 527, 1959.
 Leslie, D. C. M., *Nature*, 181, 403, 1958.
 O'Keefe, J. A., A. Eckels, and R. K. Squires, Shape of the earth from the analysis of Vanguard data, *U. S. Natl. Aeronaut. & Space Administration Publ.*, 1959.
 Parkyn, D. G., *Nature*, 182, 787, 1958a.
 Parkyn, D. G., *Am. J. Phys.*, 26, 436, 1958b.
 Parkyn, D. G., *Nature*, 183, 1045, 1959.
 Schilling, G. F., and C. A. Whitney, *Smithsonian Astrophys. Observatory Spec. Rept.* 18, 13, 1958.
 Scott, J. M. C., *Nature*, 180, 1467, 1957.
 Sterne, T. E., *Science*, 127, 1245, 1958.
 Veis, G., *Smithsonian Astrophys. Observatory Spec. Rept.* 23, 1, 1959.

(Manuscript received July 28, 1959; revised October 7, 1959.)

On the Long-Term Variation in the Cosmic Radiation¹

J. A. LOCKWOOD

*Department of Physics
University of New Hampshire
Durham, New Hampshire*

Abstract. From an analysis of the cosmic-ray intensities recorded at Mt. Washington, Mt. Norikura, and Huancayo from 1954 to 1958, the long-term variation of intensity is interpreted as occurring in several sudden decreases rather than gradually. The largest changes of cosmic-ray intensity and in the ratios of intensities recorded by the three detectors followed the occurrence of large Forbush-type decreases. The depressed intensity following the rapid decrease of July 11, 1959, further supports this interpretation. Frequency distributions of the changes in nucleonic intensity from day to day show the importance of the Forbush decreases to the long-term variation. These results are discussed in terms of possible solar modulating mechanisms.

Introduction. The results of *Forbush* [1954, 1958], *Meyer and Simpson* [1955], *Fenton and others* [1958], *Neher and Anderson* [1958], and *Lockwood* [1958] clearly establish the inverse relationship between the intensity of cosmic radiation measured on the surface of the earth and the cycle of solar activity. The magnitude of the decrease in nucleonic intensity recorded at Mt. Washington from July 1954 to March 1958, near the maximum in solar activity, was 25 per cent. A comparable decrease was recorded at Ottawa, Canada [*Fenton and others*, 1958]. The Chicago group [*Simpson*, 1958] reported essentially the same decline of the nucleonic intensity. There is no doubt about the worldwide nature and magnitude of the decrease. It is also definitely established that the magnitude of the decrease depends inversely upon the energy of the primary radiation [*Fenton and others*, 1958].

There is, however, some question about the manner in which the decreased intensity, now prevalent, has arisen. *Fenton and others* [1958] concluded that there was a gradual decrease of cosmic-ray intensity with increasing solar activity, and that there were large, transient decreases superimposed upon this slow decline. On this basis, they have corrected for the effect

of the transient decreases and found that there was an over-all gradual decrease of 20 instead of 22 per cent in the nucleonic intensity at Ottawa from 1954 to 1957. From the observed change in the nucleonic intensity at Mt. Washington, however, *Lockwood* [1958] concluded that the intensity decrease from 1954 to 1958 came about through a series of large and sudden drops, from which only partial recovery occurred. This analysis was based upon the assumption that the effect of the transient decreases cannot be eliminated in the studying of the long-term variation, since, as a result of these decreases, the largest changes in over-all intensity occur. This does not mean that the lower intensity prevailing at the maximum of the solar activity cycle is due to the recurrence of large numbers of symmetrical short-term decreases, which reduce the average for any month.

We have made a further analysis of this long-term variation to substantiate our original tentative conclusions [*Lockwood*, 1958], which were different from those of *Fenton and others* [1958]. First, the nucleonic intensities recorded by standard neutron monitors at Mt. Washington and Mt. Norikura were compared with the total intensity recorded by a shielded ionization chamber at Huancayo [*Forbush*, 1958]. The cutoff rigidity for Mt. Washington is ~ 1.0 beV/c, for Mt. Norikura ~ 10 beV/c, and for Huancayo ~ 15 beV/c. It should also be noted

¹Supported by the Geophysical Research Directorate of Air Force Cambridge Research and Development Command, and U. S. National Committee for the IGY

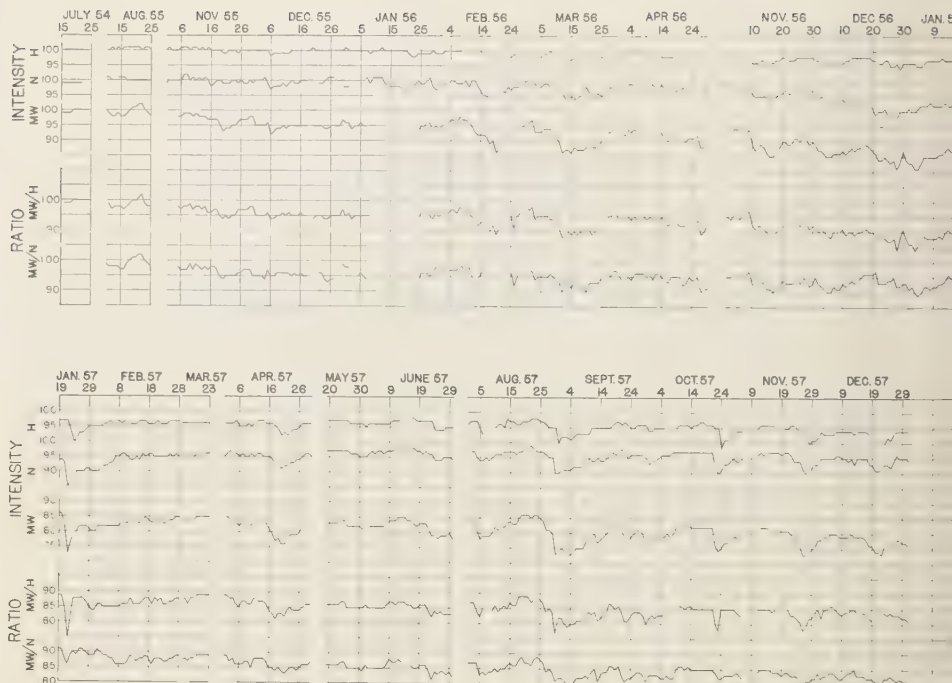


Fig. 1. The cosmic-ray intensity at Mt. Washington, Mt. Norikura, and Huancayo from 1954 to 1958. The data were obtained at Mt. Washington and Mt. Norikura with nucleonic detectors and at Huancayo with a shielded ionization detector. The intensities are normalized to the bimonthly average for July-August, 1951. The ratio of the intensity at Mt. Washington to Mt. Norikura (MW/N) and the ratio of the intensity at Mt. Washington to Huancayo (MW/H) are also plotted. Where there are gaps in the data plot, the intensity remained essentially constant.

that the maximum in the differential response curve for a shielded ion chamber occurs at about 15 beV/c, whereas for a nucleonic detector it is at about 5 to 6 beV/c. Hence, the variations observed at Huancayo are due to primaries with momenta >15 beV/c; at Mt. Norikura, to primaries with momenta >10 beV/c; and at Mt. Washington, to primaries with momenta >1 beV/c. At the last station the greatest response is to primaries with momenta 5 to 6 beV/c. If the long-term variation occurs primarily as the result of certain Forbush decreases, the intensity levels and the ratios of the intensities recorded at these stations should be changed after these events. Until another Forbush decrease occurs, these quantities should remain essentially constant, perhaps varying about some lower value or tending to revert

toward the corresponding values for 1954. Second, the nucleonic intensity at Mt. Washington for the period following the Forbush events of July 11-18, 1959, was compared with the intensity during the earlier part of the solar cycle. Third, the shielded ionization chamber data from Huancayo for the previous solar cycle were studied. Fourth, frequency distributions were made of the changes of nucleonic intensity from day to day at Mt. Washington.

Results. First, in Figure 1 are reproduced portions of the intensity data recorded at these stations from 1954 to 1958. Where there are gaps in the data plotted, the intensity remained essentially constant. Hereafter, we shall refer to the observed sudden decreases in cosmic-ray intensity, the so-called Forbush-type decreases as *F* events. No distinction will be made be-

een sharp decreases occurring in a few hours and slower ones taking place over a few days. It is seen that the intensity both at Huancayo and at Mt. Norikura was constant through 1955. At Mt. Washington it decreased slowly during November 11–19, and then decreased 12 per cent after the F event on November 19, 1955. It had not completely recovered by the highly disturbed period of February, 1956. This is noted by *Forbush* [1958], and has been shown by *Neher* [1956] and *Simpson and Meyer*, [1957] to be due to the exclusion of low-energy primary particles before the principal changes in the energy spectrum.

In February 1956, a large drop in intensity at all three stations occurred, but the intensity recovered by March 1956. In the early part of March it decreased at these same stations and the ratio of the intensities changed, indicating a change either in the primary energy spectrum or in the cutoff rigidity. It is possible that both occurred.

The intensity slowly recovered through October 1956, followed by a sudden drop on November 10, 1956, and a large change in the ratio of intensities. Several large decreases occurred through January 1957. The resulting change in intensities and ratios after the F event of January 21, 1957, were small in comparison with this decrease.

The intensity remained relatively constant, except for small changes in April and June 1957, until the large decrease of August 29, 1957. After this F event, the intensity at Mt. Washington and Huancayo was reduced and the ratio of these intensities changed. For example, the ratio of Mt. Washington to Huancayo remained below the level of the preceding period until October 1957. It is interesting to note that no changes in intensity level or ratio occurred as a result of the F event of October 1, 1957. This absence of a change in intensity level after F events often occurs and was also evident in the previous solar cycle. Small changes in the intensity and the ratio took place after the slow F event of November 22 and the sharper one in December 1957.

We do not have the data from Huancayo after December 1957, but a comparison of Mt. Washington and Mt. Norikura shows a ratio change of 2 per cent in January 1958. The

nucleonic intensity appeared to have reached a minimum in March 1958, the monthly mean intensity at Mt. Washington being 75.8 per cent of the 1954 level. Between March 1958 and July 1959 the nucleonic intensity remained depressed, slowly increasing to a level of 83 per cent by June 1959. In this period there were not as many large sudden F events as had occurred in 15 months preceding the minimum.

In Table 1 are listed some of the large F events that occurred between 1954 and 1958 and the resulting changes of the intensity level. Only those events for which the magnitude of the decrease and the rate of the decrease at Mt. Washington were greater than 5 and 1 per cent hr^{-1} respectively were considered. The periods used for determining the undisturbed or relatively stable intensity were selected according to essentially the same procedure as that of *McCracken* (paper to be published). From this table, we see that:

(a) The energy dependency of the F events was considerably greater at the beginning of the solar cycle. This has previously been reported by *Fenton and others* [1958], *Sandstrom and Forbush* [1958], and *McCracken* [1959]. It is important to note that the greatest change in the ratio occurred after the major decrease of the intensity in November 1956.

(b) The energy dependency of the long-term variation as indicated by the ratio R_s decreased, particularly after the period from November 1956 to January 1957.

Second, on July 11, 1959, the first of a series of three F events occurred. These were the largest decreases observed during this cycle of solar activity. The nucleonic intensity at Mt. Washington for these events and the succeeding months is plotted in Figure 2. For reference, the monthly mean intensity for March 1958 and the lowest value of the daily average intensity prior to these F events are indicated on the graph. The magnitude of this event can be judged by the fact that the intensity for a few hours was only 60 per cent of the 1954 intensity. Two very interesting aspects of these events are evident. First, the intensity remained depressed for about 31½ days after the first F event of July 11, 1959. Second, the intensity continued to be depressed for several months after these decreases. (Two smaller decreases

TABLE 1. Ratio of the Amplitudes of the F Events from 1954 to 1958 and the Resulting Changes in Intensity Levels

MW = Mt. Washington nucleonic detector; H = Huancayo ionization detector; MN = Mt. Norikura nucleonic detector. δ MW and δ H are the percentage decreases from the intensity at these stations during July–August, 1954.

Date of F Event	Period for Obtaining Intensity Level	$R_1 = \text{MW}/\text{H}$	$R_2 = \text{MW}/\text{MN}$	δ MW	δ H	$R_3 = \delta\text{MW}/\delta\text{H}$
11/19/55	12/12–21/55	4.5	2.2			
2/10/56		3.0	1.3	5	0	...
3/3/56		3.0	1.5			
3/12/56		4.0	1.6			
	4/ 3–12/56 } 6/22–7/1/56 }			7	1	7
9/2/56	10/16–25/56	4.0	2.0	6	1	6
11/9/56	2/16–25/57 } 3/21–30/57 }	4.0	2.0			
1/21/57		2.0	1.6	16	4	4
				15	3	5
4/17/57	5/ 7–16/57 } 7/24–8/2/57 }	2.0	2.0	17	3	5 ^a
8/29/57	10/11–20/57	1.6	1.6	18	4	4 ^b
10/21/57	11/ 9–17/57	0.9	1.0	19	4	5
11/24/57	12/ 4–13/57	1.4	1.3	21	6	3 ^b
12/19/57		1.5	1.5			

with full recovery occurred on August 20 and September 3.) The monthly mean intensity for September was 78.8 per cent and for the first half of October 80.5 per cent of the 1954 level. The ratio of the percentage change in nucleonic intensity at Mt. Washington to the meson intensity at the Massachusetts Institute of Technology detector was ~ 4.2 for the first event (McCracken, private communication). This is comparable with the ratio for the long-term variation [*Fenton and others*, 1958; *McCracken*, 1959] and for F events occurring early in the solar cycle. The continued depression of the cosmic-ray intensity after these F events is additional evidence that the greatest changes in the cosmic-ray intensity occur at the times of such events. Clearly the magnitudes of these events are not necessarily related to the magnitude of the effect upon the long-term variation.

Third, the ion-chamber data from Huancayo [*Lange and Forbush*, 1948, 1957] for the previ-

ous solar cycle show that the reduced intensity at the maximum in solar activity did not occur gradually. A large F event, starting February 3, 1946, decreased the intensity for February to 3 per cent below that of January. The intensity gradually recovered by February 1947 to within 0.5 per cent of the level of January 1946. Large F events about March 11, April 17, and May 23, 1947, from which there was only partial recovery, left the intensity in June 1947 more than 4 per cent below that of January 1946.

Fourth, a frequency distribution of the intensity changes from day to day at Mt. Washington shows the effect of these F events upon the long-term variation. Let us define $\delta I = I_{n+1} - I_n$, where I_{n+1} is the nucleonic intensity on day $n + 1$, expressed as a percentage of the average intensity for July–August, 1954, and I_n is the same quantity for day n . Then δI measures the change of intensity from day to

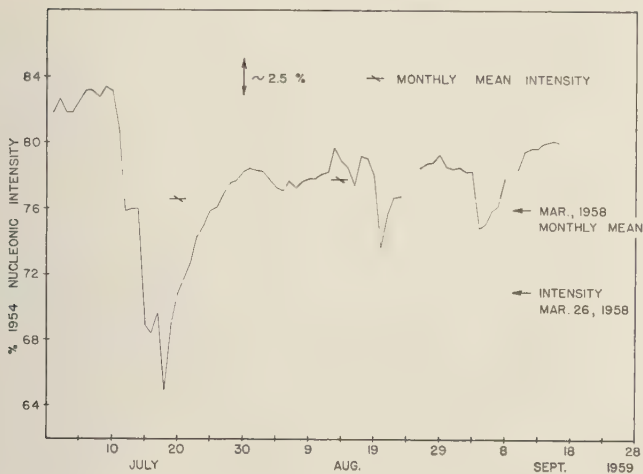


Fig. 2. Daily average nucleonic intensity at Mt. Washington, N. H. The scale at the left-hand side is the percentage intensity based upon 100 per cent for July–August, 1954.

ay, $\delta I < 0$ indicating a decrease of intensity and $\delta I > 0$ indicating an increase. In Figure 4, we have plotted a frequency distribution of I intervals of 0.2 per cent day⁻¹ for the period from July 1954 to October 1959. The frequency distribution is definitely skewed as indicated by the value of the third moment of the distribution. This is to be expected, since there were many large F events which were clearly asymmetrical; i.e., the rate of decrease was much more rapid than the rate of recovery. The mean daily rate of decrease was -0.016 ± 0.024 per cent day⁻¹. However, if these F events are detected from the data according to the criteria suggested by Fenton and others [1958], the second distribution shown in Figure 3B results. The frequency distribution curve has a smaller standard deviation, as anticipated, but the maximum of the frequency distribution remains at $\delta I = +0.2$ per cent day⁻¹ and the mean $\delta \bar{I}$ is -0.027 ± 0.023 per cent day⁻¹. If there were gradual decline of cosmic-ray intensity with superimposed transient decreases, the maximum frequency distribution should occur at some value of $\delta I < 0$. In neither Figure 3A nor 3B is the distribution curve normal, although that in 3B is closer to a normal curve as measured by the chi-square test [Mode, 1951].

Discussion. Several different solar-controlled modulating mechanisms have been proposed to

account for F events and the long-term variation. The F events may be produced by the passage of the earth into a gas cloud, such as suggested by Cocconi and others [1958], or a solar beam [Alfvén, 1946; Dorman, 1957]. The continued suppression of the galactic intensity and change in energy spectrum with associated changes in cutoff rigidities would appear to result from the effects of these same solar clouds or beams. If this gas cloud continued to expand, it would gradually fill up to the normal galactic flux over a period of, perhaps, several months. Within this large, expanded gas cloud, new clouds may originate at the sun to produce additional F events. Whether such drops would be followed by a continued reduced intensity would depend on the order and magnitude of the magnetic fields in the clouds.

On the other hand, Parker [1958] has suggested that the solar wind produces a heliocentric shell of disordered fields at a few earth-sun radii, providing a barrier through which the full galactic cosmic-ray flux must diffuse into the earth-sun region. The F events, on the basis of this model, may be caused by the interaction of the occasionally enhanced solar wind with the geomagnetic field at large distances, thus disordering the fields and affecting the arrival at the earth of cosmic-ray particles. The possibility exists that a similar grossly ordered

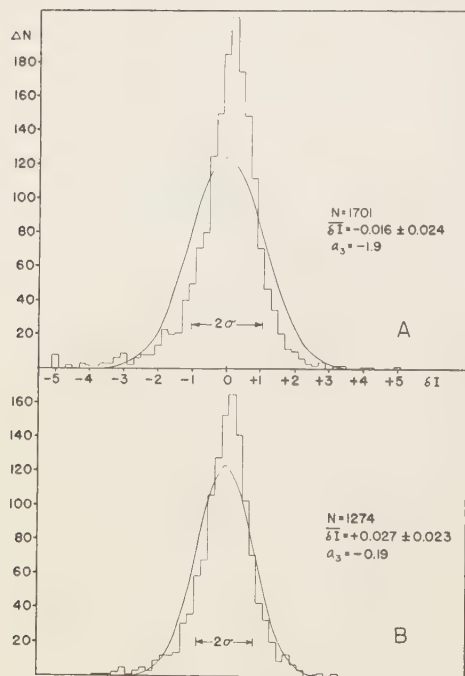


Fig. 3. Frequency distributions of the day-to-day intensity changes ΔI . $\bar{\Delta I}$ is the mean value of ΔI ; a_3 is the third moment of the distribution; and N is the total number of days. The smooth curve is the normal distribution with standard deviation σ .

A. Frequency distribution for July 1954 to October 15, 1959.

B. Frequency distribution for the same period with the transient decreases removed.

barrier would be formed by the expanding gas clouds previously mentioned.

From Figure 1, and particularly from Figure 2, it appears that the intensity decline was not a slow, gradual process. The frequency distributions of ΔI also support this interpretation. Therefore, if the heliocentric barrier proposed by Parker [1958] is to provide the long-term variation, it must be built up rapidly, within a few days after some of the F events. This seems possible on the basis of the velocity of the solar wind. Also, the modulating mechanism suggested by Cocconi and others [1958] is adequate in this respect.

The data presented in Table 1 can be inter-

preted as evidence that the long-term variation and F events are closely related. The ratios R_1 and R_2 decrease toward the maximum of solar activity in 1958, in agreement with observations of Fenton and others [1958], Sandstrom and Forbush [1958], and McCracken [1959]. The change is to be expected, since the F events late in the solar cycle occurred when the energy spectrum had been flattened and depleted of low-energy particles. From this table, it appears that the long-term variation, or, perhaps more precisely, the residual effect, of some F events became less energy-dependent toward the maximum of the solar cycle. The low-energy part of the spectrum was reduced early in the solar cycle, by about March 1956, as reported by Forbush [1958]. Since November 1956 the ratio R_3 has remained practically constant. The difference in the energy dependency of the F events and the long-term variation would be expected either on the basis of the model of Parker [1958] or that of Cocconi and others [1958], since two different effects of the same solar-controlled mechanism are being observed. The observed energy dependence of the F event arises from either the effect of the solar wind upon the geomagnetic field (Parker) or the rapid passage of the earth into a solar gas cloud (Cocconi and others). The energy spectrum of the residual effect or long-term variation is determined primarily by the energy dependence of the diffusion rate of the cosmic-ray flux through the heliocentric barrier of disordered magnetic fields formed by the instabilities in the solar wind or the expanding gas clouds. No conclusion can be drawn about the adequacy of either of these models, except that the heliocentric barrier must be formed abruptly.

In conclusion, we suggest that the intensity decrease from 1954 to 1959 was not a slow gradual process but occurred in several sudden drops (F events) from which there was only partial recovery. Consequently, the heliocentric barrier suggested to maintain the depressed intensity at the earth must develop rapidly in comparison with its decay time. The difference in the energy dependency of the F events and the long-term, or residual, variation is the result of observing two different effects of the solar wind or gas clouds.

Acknowledgments. We are grateful to Dr. Yu. Miyazaki for making available to us the electronic intensity data from Mt. Norikura. Discussions with Dr. Kenneth G. McCracken have been most helpful. The data reduction was performed by Miss M. A. Shea and Mr. Lincoln Hibbard.

REFERENCES

- Arvidén, H., *Nature*, **158**, 618, 1946.
- Becconi, Gold, Greisen, Hayakawa, and Morrison, *Suppl. Nuovo cimento*, **8**, 161, 1958.
- Chorman, L. I., *Cosmic-Ray Variations*, State Publishing House for Technical and Theoretical Literature, Moscow, 1957. (Technical Document Liaison Office, Wright-Patterson Air Force Base.)
- Fenton, A. G., K. G. Fenton, and D. C. Rose, *Can. J. Phys.*, **36**, 824, 1958.
- Forbush, S. E., Jr., *J. Geophys. Research*, **59**, 525, 1954.
- Forbush, S. E., Jr., *J. Geophys. Research*, **63**, 651, 1958.
- Lange, I., and S. E. Forbush, *Researches of the Department of Terrestrial Magnetism, Carnegie Inst. Wash. Publ.* **175**, vol. XIV, 1948; vol. XX, 1957.
- Lockwood, J. A., *Phys. Rev.*, **112**, 1750, 1958.
- McCracken, K. G., *Phys. Rev.*, **113**, 343, 1959.
- Meyer, P., and J. A. Simpson, *Phys. Rev.*, **99**, 1517, 1955.
- Mode, E. B., *Elements of Statistics*, Prentice-Hall, New York, p. 153, 1951.
- Neher, H. V., *Phys. Rev.*, **103**, 228, 1956.
- Neher, H. V., and H. Anderson, *Phys. Rev.*, **109**, 608L, 1958.
- Parker, E. N., *Phys. Rev.*, **110**, 1445, 1958.
- Sandstrom, A. E., and S. E. Forbush, Jr., *J. Geophys. Research*, **63**, 876, 1958.
- Simpson, J. A., *Bull. Intern. Geophys. Year* **15**, 11, 1958.
- Simpson, J. A., and P. Meyer, *Phys. Rev.*, **106**, 568, 1957.

(Manuscript received September 25, 1959; revised October 21, 1959.)

Decrease of Cosmic-Ray Intensity on February 11, 1958¹

J. A. LOCKWOOD

*University of New Hampshire
Durham, N. H.*

Abstract. An analysis has been made of the cosmic-ray data recorded by the IGY network of stations during the sudden decrease of intensity on February 11, 1958. The unusual rapidity of the decrease, which was followed by a temporary recovery, provides a critical test for existing solar-controlled modulating mechanisms. The main decrease, occurring practically simultaneously over the earth, was relatively energy-insensitive. The subsequent temporary recovery appeared to be associated with a spatial anisotropy. It is suggested that these variations can be interpreted as the result of disordering the outer geomagnetic field by the solar wind, as first pointed out by Parker.

I. Introduction. The sudden decrease of cosmic-ray intensity recorded by the world-wide network of IGY stations during the storm of February 10-11, 1958, was unusual in two respects: the rapidity of the decrease; and the temporary recovery of intensity following the initial decrease. It thus provides an excellent test for the various physical models proposed to explain such decreases. For example, the magnitude of the decrease recorded by the Mt. Washington nucleonic detector was ~ 5 per cent, which is not a large Forbush decrease. The rate of decrease, however, was about 7 per cent hr^{-1} , which is unusually large. Nucleonic detectors at lower geomagnetic latitudes exhibited almost complete recovery 6 hours after the initial decrease. This recovery lasted about 6 hours, depending somewhat on the location of the station.

Winckler and others [1959] have presented a very complete discussion of the solar-terrestrial phenomena related to this event. They reported the occurrence of X-ray bursts at balloon altitudes coincident with two large magnetic bays, and noted that cosmic-ray modulation was not exceptionally large, at least in comparison with the scale of the auroral and magnetic phenomena. (We shall make reference to this article for the time sequence of events during this

storm.) *Palmeira and Williams* [1958] and *Steljes and Carmichael* [1958] have reported on the rapidity of the decrease as recorded, respectively, by a meson detector of very high counting rate and a nucleonic detector. These two detectors exhibit remarkable tracking during this event. *Lockwood* [1958] also compared the intensity variation recorded by a nucleonic detector with that of the Massachusetts Institute of Technology meson detector, and found that the onset times for the main decrease were the same within the probable error of ~ 15 minutes.

A study of this cosmic-ray decrease utilizing some of the data from the IGY network has been made by *Sarabhai and Palmeira* [1959], emphasizing the anisotropy present. *Kondo and others* [1959] have investigated the world-wide pattern of the recovery following the main decrease in this event, as well as the one on September 13, 1957. They refer to this recovery as an 'increase' and conclude that the increase was due to a decrease of cutoff rigidities produced by the depression of the geomagnetic field during the severe magnetic storm.

We have made a study of this event using the data recorded by the IGY network of cosmic-ray stations. We are particularly interested in the following features of this decrease: (1) a precursory increase; (2) differences in onset times; (3) the energy dependence of the magnitude of the decrease; (4) the latitude and longitude dependence of the temporary recovery. This analysis is based principally upon the data recorded by nucleonic detectors for two

¹Supported by the Geophysical Research Directorate of the Air Force Cambridge Research and Development Command, and U. S. National Committee for the IGY.

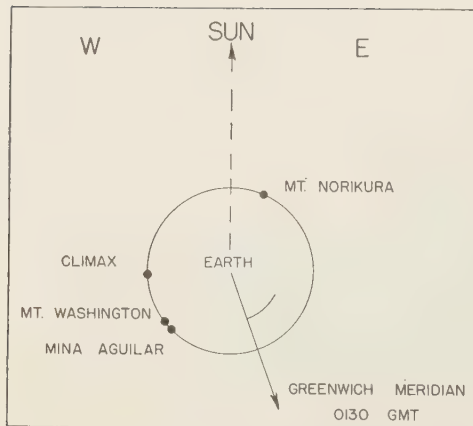


Fig. 1. Orientation of the earth with respect to the sun at 0130 UT, February 11, 1958.

reasons: the corrections for variations in atmospheric conditions are simple; and the effective primary spectrum lies in the energy range most sensitive to the changes in geomagnetic cutoff energies, owing to differences in location on the earth of the detectors.

II. Preliminary phase of the decrease. Figure 1 shows the orientation of a few cosmic-ray stations with respect to the sun at 0130 UT, February 11, 1958. The nucleonic intensities recorded by these stations in 15-minute intervals during the early phase of this decrease are plotted in Figure 2. The horizontal component of the geomagnetic field at Fredericksburg, Maryland, is also shown for comparison. From Figure 2 the following features of the initial part of this F event are evident:

(a) There is some indication of a slight increase of intensity about 0030–0100 UT at Climax, Deep River, and Mt. Washington. Such an increase could be produced by the albedo particles from a moving solar gas cloud. It is not evident at the Mina Aguilar or Mt. Norikura stations, which have much higher rigidity particles, nor is it evident for the Massachusetts Institute of Technology meson detector, which responds to higher rigidity particles.

(b) The intensity decreased ~ 1 per cent at Climax, Mt. Washington, and Deep River after 0115 UT, before the SC geomagnetic storm at 0125 UT. This feature has already been dis-

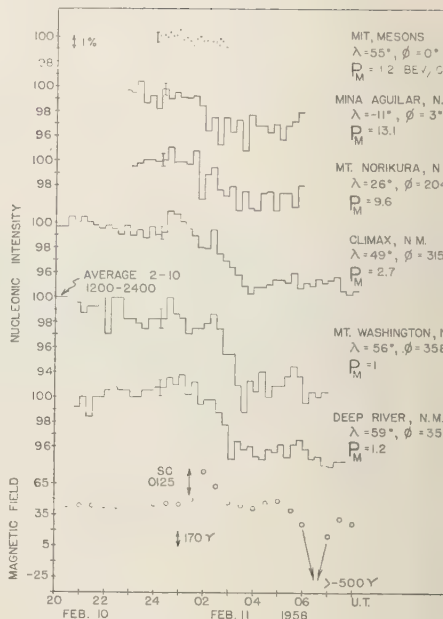


Fig. 2. Nucleonic intensity during the early phases of the decrease. The meson intensity recorded by the MIT detector and the horizontal intensity of the geomagnetic field at Fredericksburg, Md., are included for comparison.

cussed [Palmeira and Williams, 1958; Stel and Carmichael, 1958].

(c) Mt. Norikura exhibited a large decrease at 0145, shortly after the SC, followed by partial recovery to 0230. With the exception of Mina Aguilar, the other stations showed the same tendency to recover.

(d) The main decrease at Mina Aguilar started at 0200 and was very rapid.

(e) The main phase of the decrease commenced at 0245, and there does not appear to be any difference in time for this phase of the decrease to take place. It appears that the preliminary decrease at about 0115 occurred simultaneously with the onset of the magnetic storm and the main phase followed the large increase of horizontal intensity. It must be concluded that the first decrease and change in the magnetic field mark the arrival of the solar gas cloud at the earth. Since the velocity of the gas cloud was estimated as 1500 km sec^{-1} [Winckler and others, 1959] no differences in onset time

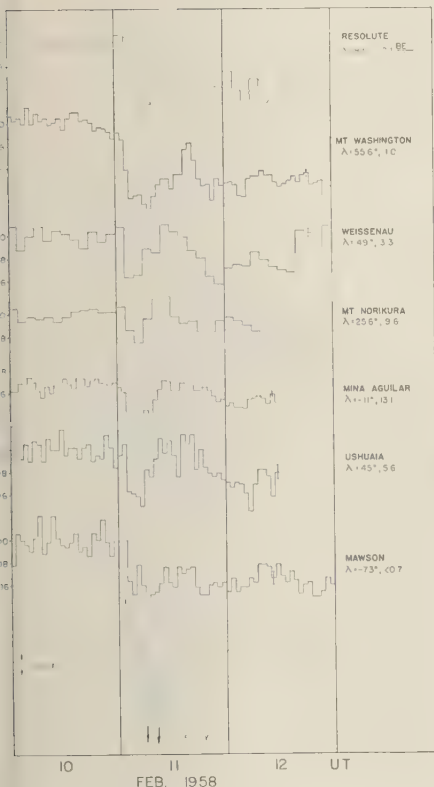


Fig. 3. Representative nucleonic intensity data, normalized to the average intensity for the period 10-24 UT, February 10. The geomagnetic latitudes are obtained from curves by McNish [1936], and the vertical cutoff rigidities, in bev/c, are the modified values calculated by Webber and Quenby [1959].

this preliminary decrease would be discernible on the time scale in Figure 2.

III. *Main phase of the decrease.* In Figure 3 are plotted representative data from the IGY network of stations to indicate the latitude dependence of the main phase of the decrease and the character of the temporary recovery between 00 and 1800 UT. The onset time for the main decrease is shortly after 0200 UT, appearing to be simultaneous at all stations to within ± 1 hour, based upon the bihourly counting rates. These plots illustrate the sharpness of the decrease. From a detailed inspection of the intensity plots, the minimum intensity was recorded at 0600 UT. The nucleonic intensity started to recover

rapidly at most stations after the first but before the second magnetic bay. After the decrease the intensity oscillated, the largest and first maximum for most stations being at 10-12 UT, with other maxima at about 18 and 24 UT, depending on the location of the station.

From the sample of data presented in Figure 3 and other data from the IGY network, we have determined the latitude dependence of the magnitude of this event; the results are given in Figure 4. The latitude dependence for stations at mountain and at sea-level elevations is shown separately, and the stations are grouped into broad longitude intervals. The magnitude of the decrease is defined as the fractional difference between the average intensity for the period 1200-2400 UT, February 10, and the lowest hourly or bihourly intensity during the main phase of the decrease. The smooth curves drawn represent the expected latitude dependence calculated from the revised differential response curves and yield functions of Webber and Quenby [1959] with an empirical modulation of the primary energy spectrum [McDonald and Webber, 1959]. The two curves labeled A are the extremes in the latitude dependence arising from the variation of cutoff rigidities with longitude. All cutoff rigidities used were those calculated by Webber and Quenby [1959]. Before interpreting the results presented in Figure 4, we shall outline the method followed to calculate the expected latitude dependence.

Webber and Quenby [1959] have shown that the differential counting rate of a nucleonic or ionizing detector located at depth x , at time t , with a vertical cutoff rigidity P can be written:

$$\left[\frac{dN(x, t)}{dP} \right]_P = \sum_z S_z(P, x) \left[\frac{dj_z(P, t)}{dP} \right]_P \quad (1)$$

where $dj_z(P, t)/dP$ is the primary rigidity spectrum and $S_z(P, x)$ is the specific yield function for the particular charge component. Thus, as Webber and Quenby [1959] showed

$$N(x, t)_P =$$

$$\sum_z \int_P^\infty S_z(P, x) \left[\frac{dj_z(P, t)}{dP} \right]_P dP \quad (2)$$

where $N(x, t)_P$ is the experimentally measured response curve of either the nucleonic or ionizing

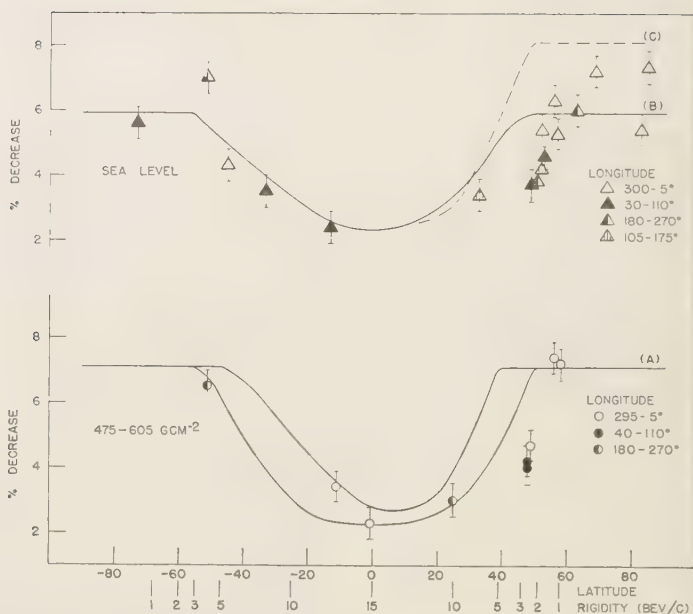


Fig. 4. Latitude dependence of the magnitude of the intensity decrease, February 11, 1958. All latitudes and longitudes are geomagnetic. The cutoff rigidities indicated are for zero geomagnetic longitude. Curve A is the calculated latitude dependence at 680 g cm^{-2} , assuming that the long-term modulation is $1 - 2.5P^{-1}$ and the superimposed transient modulation is $1 - 0.9P^{-1}$. Curve B is the same as A, but calculated for sea level (1030 g cm^{-2}). Curve C is for a modulation factor $= 1 - 0.9P^{-1}$, acting upon the 1954 primary energy spectrum.

component as determined from latitude surveys. Webber and Quenby [1959] have constructed differential response curves from the recent experimental results on the latitude dependence of these components. From these differential response curves and the yield functions, the time variations can be studied directly. We shall make the simplifying assumption that, in the latitude-sensitive part of the energy spectrum, the variation is confined to the proton component only. Hence, the differential energy spectrum at any time t can be written [McDonald and Webber, 1959]:

$$\frac{dj(P, t)}{dP} = \frac{dj(P, t_0)}{dP} [1 - K(t)P^\gamma] \quad (3)$$

where $(dj/dP)(P, t_0)$ represents the undisturbed primary spectrum. The fractional change in counting rate of a detector at depth x with cutoff rigidity P is

$$\left(\frac{\Delta N}{N}\right)_{P,x} = \frac{\int_P^\infty (dN/dP)_{P,x} K(t) P^\gamma dP}{N(P, x)} \quad (4)$$

In applying this type of modulation factor to an F event, such as the one considered here, the primary spectrum existing just before the event must be known. From the measurements of the long-term variation for nucleonic detectors at different latitudes [Simpson, 1958; Lockwood, 1958; Fenton and others, 1958; McDonald and Webber, 1959; McCracken, 1959], the modulation factor may be determined as $1 - 2.5P^{-1}$ for altitudes of 680 g cm^{-2} . The same rigidity dependence was assumed for the sea-level nucleonic detectors. There is probably a difference in rigidity dependence of the long-term variation with altitude, but the final results are not in sufficient agreement with calculations to warrant such a separation [McCracken, 1959]. Thus, just before the F event the differential

ary rigidity spectrum was

$$\frac{dj(P, t)}{dP} = \frac{dj(P, t_0)}{dP} [1 - 2.5P^{-1}] \quad (5)$$

on, the differential primary spectrum $dj(P, t)/dP$ at the minimum of the Forbush decrease was, following the method of *McDonald Webber* [1959],

$$\frac{dj(P, t_1)}{dP} = \frac{dj(P, t)}{dP} [1 - 0.9P^{-1}] \quad (6)$$

re the modulation factor assumed for the event was $1 - 0.9P^{-1}$. The latitude dependence of the F event thus calculated for sea-level and mountain elevations is plotted in Figure 4, showing that the decrease of intensity at the magnetic equator was 2 per cent. The latitude curves at the two altitudes are different because of the differential response of the two detectors depends upon altitude. These particular modulation factors appear to fit the experimental data best. For comparison, the same variation of the primary spectrum for the F event, $-0.9P^{-1}$, was applied to the primary spectrum measured in 1954, $dj(P, t_0)/dP$. The resulting latitude dependence is much steeper. We do not assume to state that these modulation factors are correct, but they appear to fit the data reasonably well.

From a more detailed inspection of Figure 4 it is evident that there was a group of stations located near latitudes 47° to 52°N for which the intensity decrease was not so large. The mountain stations varied in geomagnetic longitude from 35° to 90°E , and the sea-level ones ranged from 90° to 183°E . It does not, therefore, appear to be a regular longitude dependence. If the calculated latitude dependence curves were flattened at lower latitudes, with a more abrupt transition at about 45° , better agreement with the experimental observations could be obtained. On the other hand, it may be that the decrease varied irregularly over the earth. It is to be noted that the scatter of points was much less in the southern than in the northern hemisphere, but this was probably due to lack of data for the former.

IV. Temporary recovery. The intensity increase subsequent to the main phase of this F event will be referred to as a temporary recovery rather than as an increase. For some sta-

tions the intensity did recover to a value greater than that recorded in the hours preceding the decrease, but we wish to make a clear distinction between this variation and increases associated with the injection of particles into the earth-sun region during solar flares.

The time of the maximum recovery as a function of geomagnetic longitude for three different cutoff rigidity intervals is shown in Figure 5. There is no apparent dependence upon longitude, clearly establishing that this phenomenon was not an enhanced daily variation. Three stations recovered at quite different times, but the recovery was not complete at these locations. As a further check, we determined the average daily variations for February 1-10 and 19-28 at Huancayo, Mt. Norikura, Sulfur Mountain, and Mt. Washington stations, and these are plotted in Figure 6. The daily variation was calculated simply by averaging corresponding bihourly intensities for the two periods. The amplitude of the daily variation at Huancayo and Mt. Norikura was small, and the maxima for all stations occurred at local noon. There is some indication of an enhanced daily variation following the F event.

To illustrate the latitude dependence of the recovery, we have plotted in Figure 7 the ratio, R , of the magnitude of the recovery to that of the decrease. No separation of mountain and sea-level stations is made, except as noted, since there did not appear to be any significant difference. In the range of latitude -50° to 50° the recovery was complete. For some stations, the intensity recovery after the decrease was higher than average intensity recorded from 12 to 24 UT on February 10. The outstanding feature was the complete recovery in the specified latitude range, with an abrupt transition at about 50° latitude, or cutoff rigidity of ~ 3 bev/c. There was no longitude dependence for this transition. The polar plot in Figure 8 substantiates that the recovery was not longitude-dependent, nor did it depend on the location of the detector with respect to the earth-sun line at the time of the F event. The question remains whether this recovery was simply energy-dependent or whether some spatial anisotropy was present. By the term anisotropy, we mean here that this recovery may have been confined to primary particles of higher energies ar-

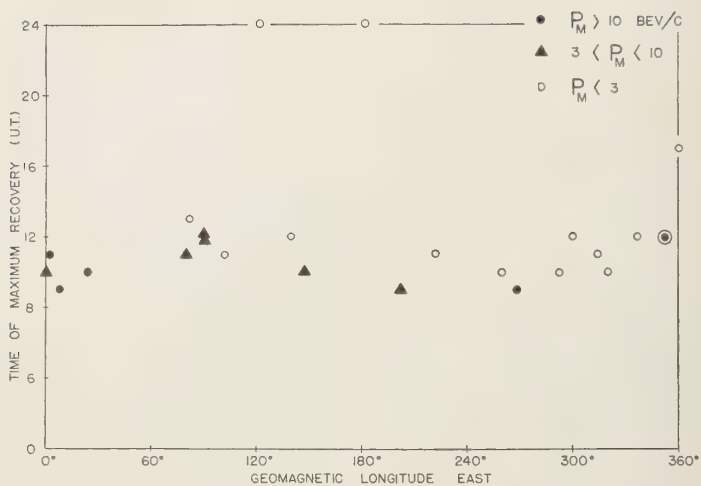


Fig. 5. The time of maximum temporary recovery from the decrease as a function of geomagnetic longitude, for stations with different cutoff rigidities.

rising at the earth from certain directions. We shall also include the possibility that these particles could arrive only in certain geomagnetic latitude ranges on the earth.

To answer this question, we have inspected the data from several standard meson telescopes, and the results are listed in Table 1. The recovery for these detectors was not complete at the higher latitudes either. We believe that the data from the Lae monitor, which were incomplete, indicated full temporary recovery. According to the calculations of *Webber and Quenby* [1959], the maxima in the differential response curves for nucleonic and mesonic detectors are at about 6 and 12 bev/c, respec-

tively. The incomplete recovery of both types of detectors at high latitudes indicates removal of protons with rigidities from the cutoff value to greater than about 12 bev/c. We must conclude, therefore, that the temporary recovery was associated with a spatial anisotropy, confined principally to a direction perpendicular to the geomagnetic equatorial plane. There is, however, an anomalous effect recorded both at the Mt. Washington nucleonic detector and at the Massachusetts Institute of Technology low energy meson detector. The two detectors tracked closely during the *F* event; both exhibited a temporary recovery about 17–18 UT, considerably later than most stations. The

TABLE 1. Temporary Recovery of Meson Intensities after *F* Event, February 10, 1958

Station	Geomagnetic		P_m , bev/c	Decrease, %	R_p
	Latitude	Longitude			
Resolute	+83°	293°E	<0.5	2.2	0.3
Churchill*	69	322	<0.5	~3.6	0.7
Sulfur Mountain	58	300	0.98	3.6	0.7
Ottawa	56	352	0.96	3.0	0.7
Lae†	-16	217	14	~2.0	~1.0
Mawson	-73	102	<0.75	2.0	<0.3

* Detector not operating throughout period.

† Operational difficulties preceding main phase of decrease.

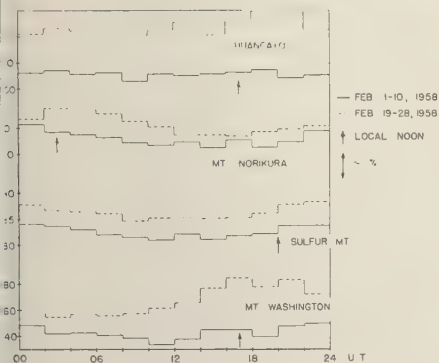


Fig. 6. Average bihourly variation of nucleonic intensity at Huancayo, Mt. Norikura, Sulfur Mountain, and Mt. Washington, before and after the decrease. Local noon is indicated by the vertical arrow.

very factor, R , was 0.85 for the meson detector and 0.73 for the nucleonic detector. This is larger than for other stations at about the same latitude. This seems to be a further indication of the occurrence of local anomalies from the general trend of this decrease.

A similar sharp decrease followed by a temporary recovery occurred on September 13, 1957 [Kondo and others, 1959]. This event is more difficult to analyze since it happened while the cosmic-ray intensity was rapidly recovering from the large Forbush decrease on August 29, 1957. Some representative data are plotted on

Figure 9. We see that the main decrease was rapid, followed by a fast recovery. At some stations the decrease was hardly evident. The increase or temporary recovery was, in general, to a lower intensity level than the level of 4 days later, as seen in Figure 9. In this event the recovery subsequent to the decrease appeared as an actual increase at some stations. The cosmic-ray flux at the earth during this event was obviously anisotropic. We shall make no attempt here to analyze this event in detail.

V. Discussion. Various models have been proposed to explain the Forbush type decreases of cosmic radiation. Singer [1958] has compared the various models, and Lockwood [1958] has discussed them in terms of the observed variations of the nucleonic component. Essentially, four different models have been suggested. Alfven [1946] showed that the decreased intensity could be produced by deceleration of the particles crossing a solar corpuscular beam. Brunberg and Dattner [1954], Venkatesan [1957], and Dorman [1957] have extended this model to fit additional experimental results. Morrison [1956] has suggested that the decrease occurs because the earth passes through a turbulent gas cloud in which the cosmic-ray flux is below the normal value. Gold [Cocconi and others, 1958] suggested that a tongue-shaped gas cloud from the sun passes over the earth, producing a decreased intensity, owing to the shielding provided by the ordered magnetic

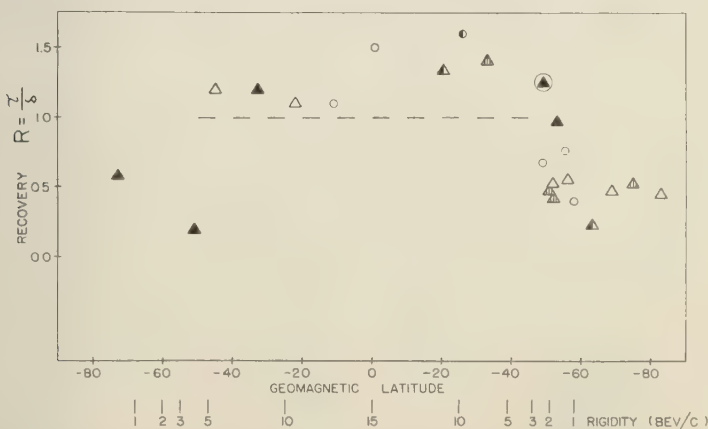


Fig. 7. The magnitude of the temporary recovery as a function of latitude. The cutoff rigidities indicated are for zero geomagnetic longitude. The longitude intervals are the same as in Figure 4.

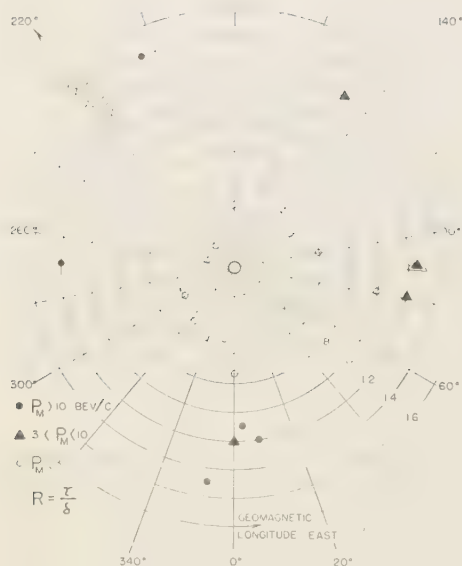


Fig. 8. The magnitude of the temporary recovery as a function of longitude for stations with different cutoff rigidities.

field in the tongue. This same field may be the guiding mechanism for low-energy solar protons to reach the earth after solar flares. *Parker* [1958] has recently shown that the solar wind may be the mechanism which produces the disturbance of the geomagnetic field and the cosmic-ray flux. In our discussion of this F event, we shall refer only to the latter two models since the rapidity of the decrease and temporary recovery do not appear to be explained by the first two.

The rapid decrease that occurred practically simultaneously over the earth is compatible with Gold's model. The tongue of ionized gas passing over the earth produced the first world-wide decrease. The temporary recovery after the main phase, which was observed to be largest at low latitudes, is more difficult to fit into this model. It does not seem reasonable that high-energy particles could be trapped in such a cloud, later to be detected primarily in an equatorial belt, ranging from 50° to -50° geomagnetic latitudes. It is possible that the cloud has a peculiar structure to provide the observed

cosmic-ray variations; we have no idea about this structure.

We believe that the observed variations during this decrease support the suggestion of *Parker* [1958] that the Forbush-type decrease is the result of disordering of the outer geomagnetic field by the solar wind. These calculations showed that the solar gas can penetrate the geomagnetic field to a considerable depth, consequently the field outside this penetration depth, R' , does not resemble a simple dip field. On quiet days the solar wind may disturb the field at distances of about 6 earth radii (R_E), corresponding to the continuous agitation of the magnetic field observed at stations above 65° N geomagnetic latitude. When the solar wind is high, the penetration depth increases $\sim 1 R_E$. *Parker* has calculated the resulting effect upon the arrival of cosmic rays at the earth, considering stations at latitudes greater than 65° . These calculations are based upon the modulation of the cosmic-ray density inside the heliocentric shell of disordered magnetic field which is different from the galactic cosmic-ray density. *Parker* [1958] attributes the long-term variation to the accumulation of such a shell, and we may assume that in 1954, at solar minimum, the cosmic-ray density in the earth's region is the full galactic density. As *Parker* points out, these calculations for the F event cannot be extended to the equatorial region without a more detailed picture of the disordering. Therefore, the complete latitude dependence of the decrease in the primary intensity, and consequently the latitude dependence at sea-level or mountain stations, cannot be evaluated.

The geocentric depression of the cosmic-ray intensity at the magnetic poles caused by the disordering of the outer geomagnetic field depends upon the scale of the disordering [*Parker* 1958]. If it is assumed that the disordering is on a small scale, the energy spectrum of the depression is steeper than for larger-scale disordering. If this spectrum could be interpreted to give the latitude dependence for the decrease it would result in a much flatter latitude curve than the empirical modulation factor of *Webb and Quenby* [1959]. The latitude dependence of the decrease shown in Figure 4 indicates that this empirical modulation factor is probably

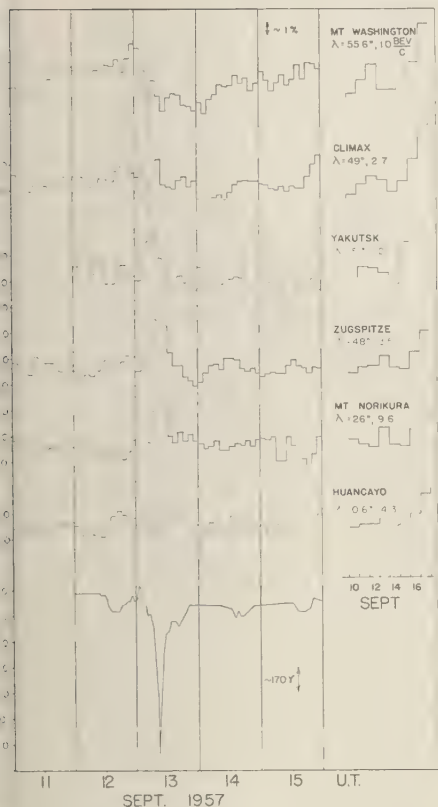


Fig. 9. Representative nucleonic intensity data during the *F* event, September 13, 1957. The daily average intensity from September 10 to 17 is shown at the right. The variation in the horizontal intensity of the magnetic field at Fredericksburg, Md., is shown for comparison.

is not completely correct. The data show that the depression was relatively energy-insensitive down to ~ 4 beV/c, then became very energy-dependent. Since the atmospheric cutoff for nucleonic detectors at sea-level or mountain elevations is 1-2 beV/c, we cannot determine the energy dependency below ~ 2 beV/c, or 52° latitude. The geocentric depression for disordering on a small scale comparable with the radius of curvature of a particle in the field [Parker, 1958] would exhibit a spectrum closer to that observed latitude dependence. It is apparent that variations in the scale of disordering would produce large changes in the energy dependence for different *F* events, which are observed.

We should like to suggest that the disordering of the outer geomagnetic field explains the variations in cosmic radiation observed during the *F* event on February 11, 1958. The sudden impact of the solar wind upon the geomagnetic field penetrating to $< 6 R_E$ provided the first main decrease, observed almost simultaneously at all stations. Both the sudden increase of the horizontal intensity of the magnetic field (see Fig. 2) and the evidence of a possible increase of cosmic-ray intensity before the main decrease, which is indicative of an albedo effect, suggest that the solar gas cloud had a sharp front. The gas density behind the front must have remained high ($> 10^8$ ions cm^{-3}) for several hours after the onset to continue to suppress the cosmic-ray intensity. The solar wind (or solar gas cloud) then decreased, disordering the geomagnetic field only at much greater distances from the earth. This agitation would maintain a reduced cosmic-ray flux at higher latitude stations ($> 50^\circ \text{N}$ or S). The temporary recovery at lower latitude stations occurred since detectors located there respond only to higher-rigidity particles (> 10 beV/c) which are little affected by the agitation at great distances. The solar wind then increased to produce the slow second decrease at equatorial stations. Since the earth's field at large distances was already disordered, producing the reduced intensity at high latitudes, we can only suggest that the relatively slow increase of the solar wind after 1200 UT, February 11, 1958, and consequent deeper penetration into the earth's field, decreased the intensity slowly, predominantly at lower latitudes.

It was noted that the meson telescopes at higher latitudes did not recover fully after the main decrease. The calculations of Parker [1958] for the effect of the disordering of the geomagnetic field upon the cosmic-ray flux measured at the earth for latitudes greater than 65° were restricted to protons with energies less than 10 beV. Since the maximum differential response of a meson telescope at sea level is at about 10 beV/c, Parker's calculations should hold approximately for the higher-energy primaries to which the meson detectors respond. The incomplete recovery of the meson detectors at high latitudes would then be expected on the basis of this model. It would also be expected

that the ratio of the intensity recovery of meson detectors to nucleonic detectors at higher latitudes would be greater than 1, as observed.

The sharp transition observed at $\sim 50^\circ$ latitude in the plot of the recovery factor, R , as a function of latitude is anticipated on the basis of such a model. Noting that nucleonic detectors deep in the atmosphere respond essentially to primary particles arriving close to the zenith, we refer to the experiments of *Brunberg* [1956] on the deflection of cosmic rays in the earth's magnetic field for an explanation. If we consider the asymptotic directions of particle orbits in the vertical plane only at great distances from the earth, we find that at higher latitudes a nucleonic detector sees particles coming from directions above (or, in the southern hemisphere, below) the equatorial plane. And it is known [*Webber and Quenby*, 1959] that at higher latitudes ($>40^\circ$) the particles tend to follow the field lines. Consequently, the continuous disordering of the geomagnetic field at greater than $6 R_E$ would have little effect on the cosmic-ray intensity at lower latitudes.

It seems reasonable that the F event of September 13, 1957, was also caused by agitation of the outer geomagnetic field by the solar wind. In Figure 9 it is seen that the temporary recovery or increase coincides with the main decrease in the horizontal component of the geomagnetic field. During the F event of February 1958, the temporary recovery, however, commenced between the times of the two main decreases of the magnetic field. It appears that the magnetic disturbances during these two events were not the same. Any detailed relationship between the geomagnetic field and cosmic-ray variations is lacking. In the large F events, a sudden commencement geomagnetic storm usually precedes or coincides with the onset of the cosmic-ray decrease [*Lockwood*, 1958]. If disordering of the outer geomagnetic field accounted for the variations during the September 1957 F event, the disordering must have been much more irregular than in the February event. This might be expected since the September F event was about 2 weeks after one of the largest solar disturbances of the present solar cycle, and the electromagnetic state of the earth-sun region was not in equilibrium. The latitude and longitude dependence

of the intensity variations, as can be seen in Figure 9, are not nearly so regular as for the F event in February 1958. *Kondo and others* [1959] have concluded that the increase observed at this time (and in February 1958) was due to a decrease of cutoff rigidities caused by the depression of the geomagnetic field in intensity during the severe magnetic storm. In other Forbush events, large geomagnetic field decreases with the same phase but no increase of intensity were observed. The lack of correspondence between magnetic field variations at the earth and cosmic-ray intensity changes renders such an interpretation difficult. The magnetic effect to be measured at the earth from disordering the outer geomagnetic field is not easily evaluated, and the lack of correspondence, therefore, is expected.

In conclusion, we suggest that the cosmic-ray decrease and subsequent temporary recovery for these F events were caused by disordering of the outer geomagnetic field by the increase of the solar wind. As a result of the penetration of the geomagnetic field by clouds of ionized gas, the field at large distances will no longer be that of a dipole. The unusual intensity variations observed during the F event of February 1958 are suggestive of such a model. We are now studying many other large Forbush decreases observed in the present solar cycle, particularly the onset times, energy dependences, and evidence for a precursory increase, since these factors are most important to determination of the correct physical interpretation of the decreases. Whether these other decreases can be explained on the same basis is not known, since, as *Parker* [1958] has pointed out, it is not possible to calculate the decreases expected at low latitudes from our present knowledge of the disordering process. Where such calculations can be made, the empirical modulation factors assumed can be verified.

Acknowledgments. The assistance of Messrs Charles Boghosian and Lincoln Hubbard in analyzing and plotting the data was most helpful. The author gratefully acknowledges the cosmic ray data supplied by Dr. John Simpson, University of Chicago (Chicago, Climax, Huancayo); Dr. Hugh Carmichael (Deep River); Mr. Ricardo Palmeira, Massachusetts Institute of Technology (meson detector); Dr. D. C. Rose, National Research Council, Ottawa (Resolute, Churchill, Sul

Mountain, and Ottawa); Dr. R. B. Brode, University of California (Hawaii and Berkeley); Yukio Miyazaki, Tokyo, Japan (Mt. Norinara); Dr. D. M. Thomson, Makerere College (Makerere); Dr. H. Ghielemetti, Buenos Aires, Argentina (Mina Aguilar, Buenos Aires, and Chuquibambilla); Dr. G. Schwachheim, Brazil (Rio de Janeiro); Dr. A. M. van Wijk, Magnetic Observatory, Hermanus (Hermanus); Dr. A. G. Fenton, University of Tasmania (Hobart [Wellington], Launceston, and Lae); Dr. V. Petržílka, Czechoslovakia Academy of Science (Lomnický štít). Without these data, and those from the IGY World Data Center A, this study would not have been possible. The magnetic data from the Fredericksburg Observatory, U. S. Coast and Geodetic Survey, were most helpful.

REFERENCES

Alfvén, H., *Nature*, 158, 618, 1946.
 Brunberg, E. A., *Tellus*, 8, 216, 1956.
 Brunberg, E. A., and A. Dattner, *Tellus*, 6, 73, 254, 1954.
 Cecconi, Gold, Greisen, Hayakawa, and Morrison, *Nuovo cimento Suppl.*, 10, 161, 1958.
 Gorman, L. I., *Cosmic-ray Variations*, State Publishing House for Technical and Theoretical Literature, Moscow, 1957. (Technical Document Liaison Office, Wright Patterson Air Force Base.)
 Fenton, A. G., K. B. Fenton, and D. C. Rose, *Can. J. Phys.*, 36, 824, 1958.
 Honda, Nagashima, and Yoshida, to be published, 1959.
 Lockwood, J. A., *Phys. Rev.*, 112, 1750, 1958.
 McCracken, K. G., *Phys. Rev.*, 113, 343, 1959.

McDonald, F. B., and W. R. Webber, *Phys. Rev.*, 115, 194, 1959.
 McNish, A. G., *Terrestrial Magnetism and Atmospheric Elec.*, 41, 37, 1936.
 Morrison, P., *Phys. Rev.*, 101, 1397, 1956.
 Palmeira, R., and R. W. Williams, *Nuovo cimento*, 8, 352, 1958.
 Parker, E. N., *Phys. Rev.*, 110, 1445, 1958.
 Sarabhai, V., and R. Palmeira, IUPAP Cosmic-Ray Congress, Moscow, 1959.
 Singer, S. F., *Progress in Elementary Particle Cosmic-Ray Physics*, Interscience Publishers, New York, pp. 295 ff, 1958.
 Simpson, J. A., *Bull. Intern. Geophys. Year*, 15, 11, 1958.
 Steljes, J. F., and H. Carmichael, *Nuovo cimento*, 8, 393, 1958.
 Venkatesan, D., *Tellus*, 9, 209, 1957.
 Webber, W. R., and J. J. Quenby, *Phil. Mag.*, 37, 90, 1959.
 Webber, W. R., and J. J. Quenby, *Phil. Mag.*, 41, 654, 1959.
 Winckler, J. R., L. E. Peterson, R. Hoffman, and R. Arnoldy, *J. Geophys. Research*, 64, 597, 1959.
 IGY Data Center A for Cosmic-Rays—Neutron Monitor Data furnished from the following stations and investigators: Fairbanks, Alaska, U.S.A., Dr. S. A. Korff; Uppsala and Murchison Bay, Sweden, Dr. A. E. Sandström; London, United Kingdom, Dr. P. Rothwell; Moscow, Yakutsk, and Alma Ata, USSR; Weissenau and Zugspitze, German Fed. Republic, Professor A. Ehmert and Dr. Heisenberg.

(Manuscript received October 5, 1959;
 revised November 5, 1959.)

Geomagnetic Effects on Cosmic Radiation for Observation Points Above the Earth

JOSEPH E. KASPER¹

State University of Iowa, Iowa City, Iowa

Abstract. The theory of geomagnetic effects on cosmic radiation is discussed for the case in which the point of observation is above the surface of the earth. Parts of the older geomagnetic theory are readily modifiable for application under the new conditions; the manner in which ordinary Störmer theory and the Lemaitre-Vallarta theory of the main cone can be adapted is discussed. However, the simple shadow cones of Schremp appear now in a new light; as the point of observation rises from the earth, the simple shadow effect which exists at the earth makes a smooth transition to a new kind of earth's shadow effect. Associated with this effect are principal shadow cones which are defined by reference to the properties of certain families of trajectories. For observation points above the earth a modification in our view of the penumbral region is required; this is discussed very briefly. Extensive computations of principal shadow cones have been carried out by machine integration of the equations of motion and subsequent abstraction of data from these trajectories. Sample computed cones are given for representative particle rigidities, geomagnetic latitudes, and altitudes of observation. The centered dipole approximation for the magnetic field of the earth is used throughout.

INTRODUCTION

A new era in the study of cosmic radiation has begun, during which direct observations of the radiation can be made at great altitudes above the earth. However, geomagnetic theory, developed and employed as an indispensable adjunct to investigations of the radiation and the earth's field during the last few decades, restricted to the case in which the observation point is on the surface of the earth. The present paper has as its purpose extension of the theory to make it applicable to points of observation at nonzero altitudes, and to present some results of numerical calculations designed to make the theory applicable to interpretation of empirical data.

It is felt at the outset that the reader must be made aware of some compromises made in this paper, these having been dictated by practical considerations related to the publication of such an elaborate piece of work. (a) Familiarity on the part of the reader with the older geomagnetic theory will be assumed. In particular, familiarity will be assumed with the manner of treatment of the problem of the reduced motion of charged particles in the meridian plane of a magnetic

dipole, as carried out by various authors [see, for example, Störmer, 1955; Vallarta, 1938], and with the modes of definition and computation of the Störmer cone [Störmer, 1955; Vallarta, 1938], main cone [Lemaitre and Vallarta, 1936a, b] simple shadow cone [Schremp, 1938a, b] and penumbral region [Hutner, 1939a, b].

(b) For the most part only newly introduced terms and symbols will be defined. The extensive symbolism and terminology of the authors cited above will be used without redefinition, when this is consistent among them, except that the symbol γ_1 commonly used in the literature will be replaced by γ for simplicity. (c) The subject treated requires very lengthy discussion when its intricacies and ramifications are to be dealt with fully, and when suitable rigor is sought. For the sake of brevity the account given here is condensed, but it is hoped that omissions of details will not hinder understanding by any reader moderately versed in geomagnetic theory. A strong effort to make this hope realistic has been made by keeping the discussion rather graphic, and by incorporating many illustrative figures. (d) It is practical to give here only a selection of the numerically computed principal shadow cones. More extensive results and more complete discussion of the theory may be found in the author's doctoral thesis,

¹ Present address: Physics Department, Coe College, Cedar Rapids, Iowa.

State University of Iowa, Iowa City, Iowa (1958), of which the work described in this paper formed a fractional part.

PRELIMINARY CONSIDERATIONS

The magnetic field of the earth will be taken to be that of a magnetic dipole lying at the geometrical center of a spherical earth. (Effects of the eccentricity of the dipole, oblateness of the earth, and higher multipole moments can be expected to be negligibly small for cosmic-ray—that is, high-rigidity—particles at high altitudes.) The numerical value of the dipole moment will be absorbed in the Störmer unit as is usual, and need not be specified.

Regions of allowed and forbidden directions of arrival in the sphere of directions at the observation point will be represented in cone diagrams of the traditional kind [Lemaître and Vallarta, 1936a, b], with coordinates $\sin \theta$, $\cos \theta \sin \eta$. Each cone diagram will now, as in the older theory, pertain to one particle rigidity and one geomagnetic latitude of observation, but will also now pertain to one altitude of observation. The geomagnetic latitude will be represented by λ_s . Particle rigidity will be specified in terms of the quantity r_s (e for 'earth'), which also measures the radius of the earth in Störmer units. The altitude of observation will sometimes be given by the ratio r_s/r_e , where r_s (s for 'satellite') is the radial distance from the center of the earth in Störmer units to the point of observation; sometimes in terms of the ratio ρ_s/ρ_e , where ρ_s and ρ_e have the same significance as r_s and r_e , respectively, but are measured in ordinary units of length; sometimes in terms of the quantity $(x_s - x_e)$, where x_s is defined by $\exp x_s = 2\gamma r_s$ and x_e by $\exp x_e = 2\gamma r_e$, in accordance with the Störmer-instituted use of the Goursat transformation; and sometimes as distances h above the surface of the earth in ordinary units of length, the units to be specified.

A cone diagram, regarded as a coordinate system, is now determined by the values of r_s and λ_s , the coordinates of the observation point in the meridian plane of the earth, and not by the values of r_e and λ_e , as in the older theory. Thus a coordinate $\sin \theta$ is determined by the formula

$$\sin \theta = (2\gamma/r_s \cos \lambda_s) - (\cos \lambda_s/r_s^2)$$

and the other coordinate $\cos \theta \sin \eta$ is then

determined by the value of θ implied by $\sin \theta$ and by the value of η for the particular trajectory in question at the point of observation.

Some bounds on the ranges of values of r_s and λ_s and parameters figuring in the problem will be in effect hereafter. (a) Values of r_s greater than 0.8 will not be considered. This will delimit the amount of computational work to be done, but is fully acceptable because it corresponds to an energy of about 37 bev for protons, that only the small high-energy tail of the primary cosmic-ray spectrum will be neglected, and because for larger rigidities the earth's geometrical shadow is often, to sufficiently good approximation, equal to the size of the corresponding principal shadow region. (b) Only positive values of γ will be considered. It can be easily verified that this restriction means that for some values of r_s , λ_s , and r_s/r_e some westerly portions of some cone diagrams will be neglected, and that this is of no practical importance. (c) Altitudes above the earth ranging from $h = 0$ to about $h = 8000$ miles above the earth's surface must be considered in order that a suitably wide range of practical interest will be dealt with. This means that r_s/r_e (or ρ_s/ρ_e) must range from 1 to about 1.0001, or that $(x_s - x_e)$ must range from 0 to about 0.0001.

APPLICABILITY OF PARTS OF THE OLDER THEORY IN THE NEW CONTEXT

The part of the meridian plane in which real solutions of the equations of motion can exist is a simply connected region when $\gamma < 1$, but when $\gamma > 1$ this part of the plane consists of two regions separated from each other by a region in which real solutions cannot exist, as is well known [Störmer, 1955; Vallarta, 1936]. Also, $r < 1$ in the inner of the two regions and $r > 1$ in the outer, when $\gamma > 1$. It follows that trajectories from infinity cannot reach an observation point for which $r < 1$ when $\gamma > 1$. This is, in brief, a summary of Störmer's theory for observation points on the earth's surface.

If $r_s \leq 0.8$ and $\gamma > 1$ (r_s/r_e may be as large as 3), the surface of the earth will always lie in the inner of the two regions of the meridian plane in which motions are allowed, and the observation point may lie either in that same part of the plane or in the outer region, depending on the value of r_s/r_e . The case $r_s < 1$, $r_s <$

$\gamma > 1$ is clearly such that Störmer theory applies. Specifically, all directions of arrival at the observation point are forbidden for particles coming from infinity. However, the case $\gamma > 1, r_0 < 1, \gamma > 1$ is very different. If the observation point lies in the outer region while the surface of the earth lies in the inner region, in all directions of arrival at the observation point for the value of γ in question are allowed, rather than forbidden, directions of arrival. This is easily proved by considering the trajectories issuing from an observation point in the outer region of the x, λ plane, remembering that particle motions in these coordinates are reversible. In that region the x component of the force is everywhere directed in the sense of increasing x . Therefore any trajectory which comes from the observation point in the plane never issues initially to the right in x and converges to infinity without any extremum in x , or else issues initially to the left in x , attains a minimum in x on or before reaching the curve $\lambda = 0$, which bounds the outer region of the meridional plane, and then travels to the right to infinity without any further extrema in x .

We turn now to the question of the main theorem. This is defined [Lemaitre and Vallarta, 1936a, b] in such a manner that the existence of an impenetrable earth plays no part in the definition; the radius of the earth figures in the theory only in fixing the radial distance of the observation point from the center of the earth, and the earth itself as an absorbing body is ignored. If any trajectory is physically forbidden for the reason that a particle following it would have to traverse part of the earth, that fact is taken account of in the theories of the simple shadow and penumbral effects. Clearly, a similar theorem can be used when the point of observation lies above the earth's surface. That is, the main cone can be defined as before, and the question of the intersection of trajectories with the impenetrable earth can be dealt with in other parts of the theory. As a consequence of the adoption of this view, the theory of the main cone as given by Lemaitre and Vallarta [1936a, b] now applies. Furthermore, the computed cones given by those authors can be interpreted as applying to points at arbitrary altitudes above the earth by substituting our r_0 and λ_0 for their r and λ . (Unfortunately, the published main cones lie only in the observer's upper hemisphere

of directions of arrival, and the ranges of r and λ are rather restricted.)

THE PRINCIPAL SHADOW CONES

Considerations preceding definition of principal shadow cones. It is now necessary to consider the effect of the solid earth as a radiation-blocking body which delimits the possible directions of arrival of cosmic radiation at the observer's position. For points on the earth, some continua of directions of arrival rendered forbidden by the earth are bounded by trajectory segments of uncomplicated forms which generate the simple shadow cones of Schremp [1938a, b], and other earth-blocked trajectories lie in the penumbral dark regions [Lemaitre and Vallarta, 1936a, b; Schremp, 1938a, b; Hutner, 1939a, b], bounded by cones generated by trajectories of more complicated kinds. One would, after a little reflection, expect that for points above the earth's surface there would be a large dark region representing a shrunken simple shadow region, and penumbral dark regions corresponding to the points for zero altitude. In each case the dark regions ought to be bounded by cones generated by grazing trajectories that at least somewhat resemble trajectories of the second kind (the generators of the simple shadow cones) and the penumbral grazing trajectories of Schremp and Hutner.

However, even a brief inspection of computed particle trajectories shows that often the bounding cones of the new 'simple shadow' regions must be generated by trajectories of more complex form than for points on the earth and that, for observation points far above the earth, the simplicity of the old theories of the simple shadow cone and the penumbral cones is lost. In this connection Figures 1, 2, and 3 may be consulted. These show some machine-computed members of pencils of trajectories radiating from two differently located points in the x, λ plane for different values of γ . Examination of the trajectories shows the presence of dark ranges of the angle η at the observation point; these ranges are bounded by values of η corresponding to trajectories with points of tangency at the earth's surface, which may be imagined to lie at will to the left of the observation points. It is apparent that such grazing trajectories may be used to define the dark regions, but it is also evident that such bounding

grazing trajectories will often be complicated in form. For example, highly oscillating specimens may be found in the figures. Too, not all grazing trajectories can form suitable bounds of dark regions, since some can be found which clearly are in the midst of continua of trajectories which intersect the earth.

Our purpose now is to render more precise these and other more extensive and general observations. Consider any trajectory which issues from a point x_0, λ_0 and reaches a point of tangency with the earth at $x = x_0$ without having intersected the earth in the interval. The segment between the initial point and the point of tangency will be called a grazing trajectory segment, abbreviated by GTS. A distinction between two kinds of GTS will now be made for the purpose of eliminating from further consideration those of one kind. (a) One can show that if the point of tangency of a GTS at the earth lies north² of the intersection of the line $x = x_0$ with the locus of minima of self-reversing trajectories,³ and if the GTS is directed northward at that point, then the continuation of the GTS past the point of tangency turns in a counterclockwise sense and cuts through the surface of the earth. It is appropriate to call the minimum in x at the earth of such a GTS a relative minimum in x . (b) One can also show that any other GTS is such that its continuation past its tangency with the earth will not intersect the earth at all, or else must cross the equator at least twice before reaching the earth again in the same hemisphere of latitudes. The proof which is based on Störmer's theorem [1955, p. 241] is easy but is not susceptible to a brief statement; it will be omitted here. In the cases of such GTS the minima in x at the points of tangency with the earth will be called absolute. This modifier means that such a continued GTS has no x minimum to the left of the earth at all, or else has one only after a long and convoluted traversal of the flat plain in the x, λ plane, including a crossing of the equator.

Only GTS with absolute minima in x at the

earth will be of interest here; each may bound a continuum of forbidden trajectories on one side whereas GTS with relative minima in x at the earth lie embedded in continua of forbidden trajectories and are themselves forbidden. To simplify the following discussions and to agree with established precedents in the nomenclature of the older theories, GTS with absolute minima at the earth will be called trajectories of the fourth kind, abbreviated T4.

At a point of observation at the surface of the earth, let η_N denote the value of η which corresponds to a trajectory of the second kind bounding a simple shadow region in its northern reach, and η_S the similar quantity for the southern reach of a simple shadow region at the point. If at the point there is no southern reach of a simple shadow region, we will take $\eta_S = 270^\circ$, and if there is no northern reach, $\eta_N = 90^\circ$. (For any point on the earth, at least the entire lower hemisphere of directions of arrival is forbidden.) Now at a point x_0, λ_0 and for a fixed value of γ , let η begin at the value η_N and increase until the value η_S is reached. Then every value taken on by η will correspond to a GTS with an absolute minimum in x at some $x_0(x_0 - x_0)$, and hence will be a T4.

It is claimed that, for at least some points x_0, λ_0 and for some values of γ , the location in x of the absolute x minima of the T4 generated by the variation of η from η_N to η_S recedes monotonically to the left in x as η increases from η_N until for some value of η the trajectory has its x minimum farther to the left in x than does any other preceding trajectory, and after this the x minima move monotonically to the right as η continues to increase to η_S . A point of this character will be called a point of kind 1, abbreviated P1. A proof of the existence of P1's will not be discussed here; however, the reader can advantageously examine Figures 1, 2, and 3 to convince himself of their existence. In the figures the movements of the x minima can be traced. (The claim of the existence of P1's ought to be attributed strictly, to be attributed to Störmer [1955, pp. 250, 253].)

For some points x_0, λ_0 and some values of γ , as η varies from η_N to η_S , the movement of the absolute x minima of the T4 consists of a monotonic recession to the left, followed by a monotonic movement to the right, followed by a second recession to the left, with a final mov-

² Some terminology used here is adapted to the case in which the point of tangency lies in the northern geomagnetic hemisphere; obvious modification can be made for application to points in the southern hemisphere.

³ Such loci are discussed, and computed examples are given, in part II of Schremp [1938b].

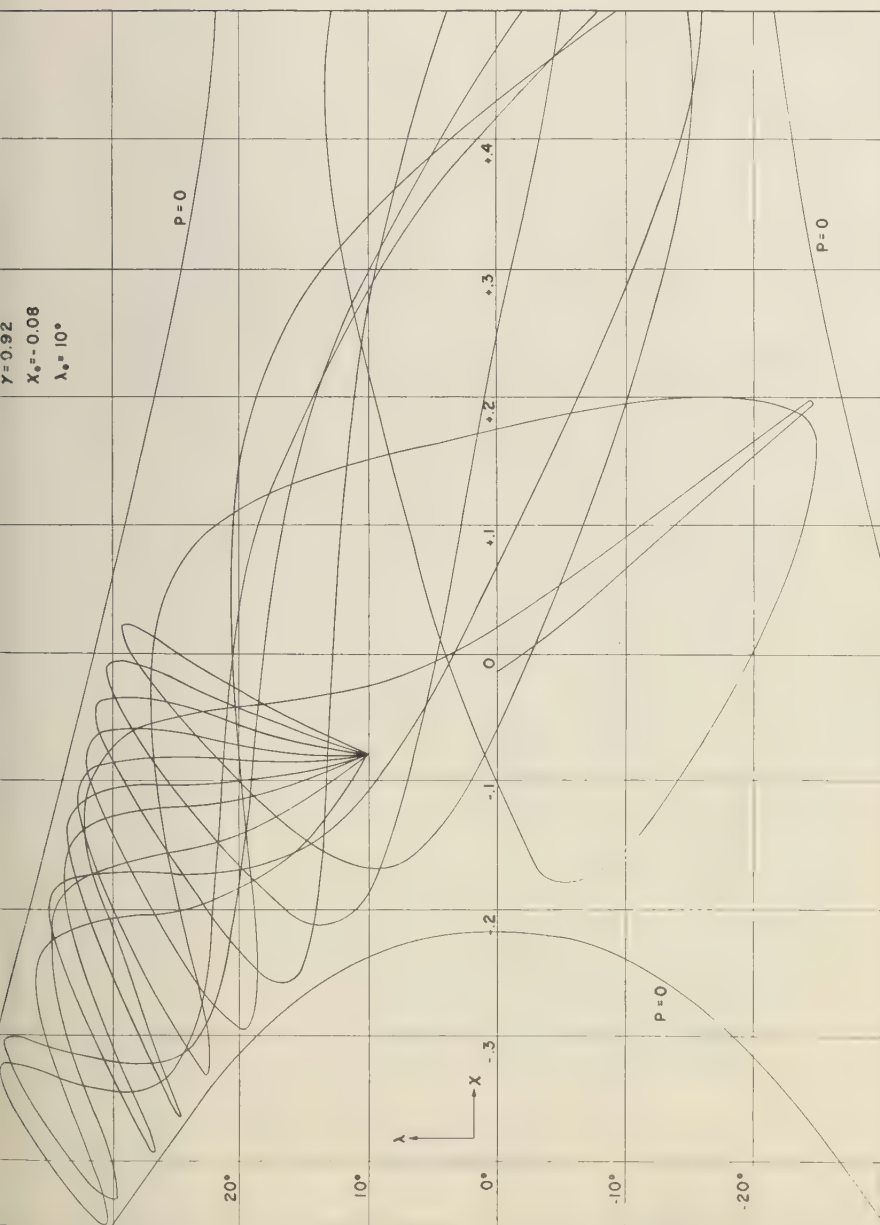
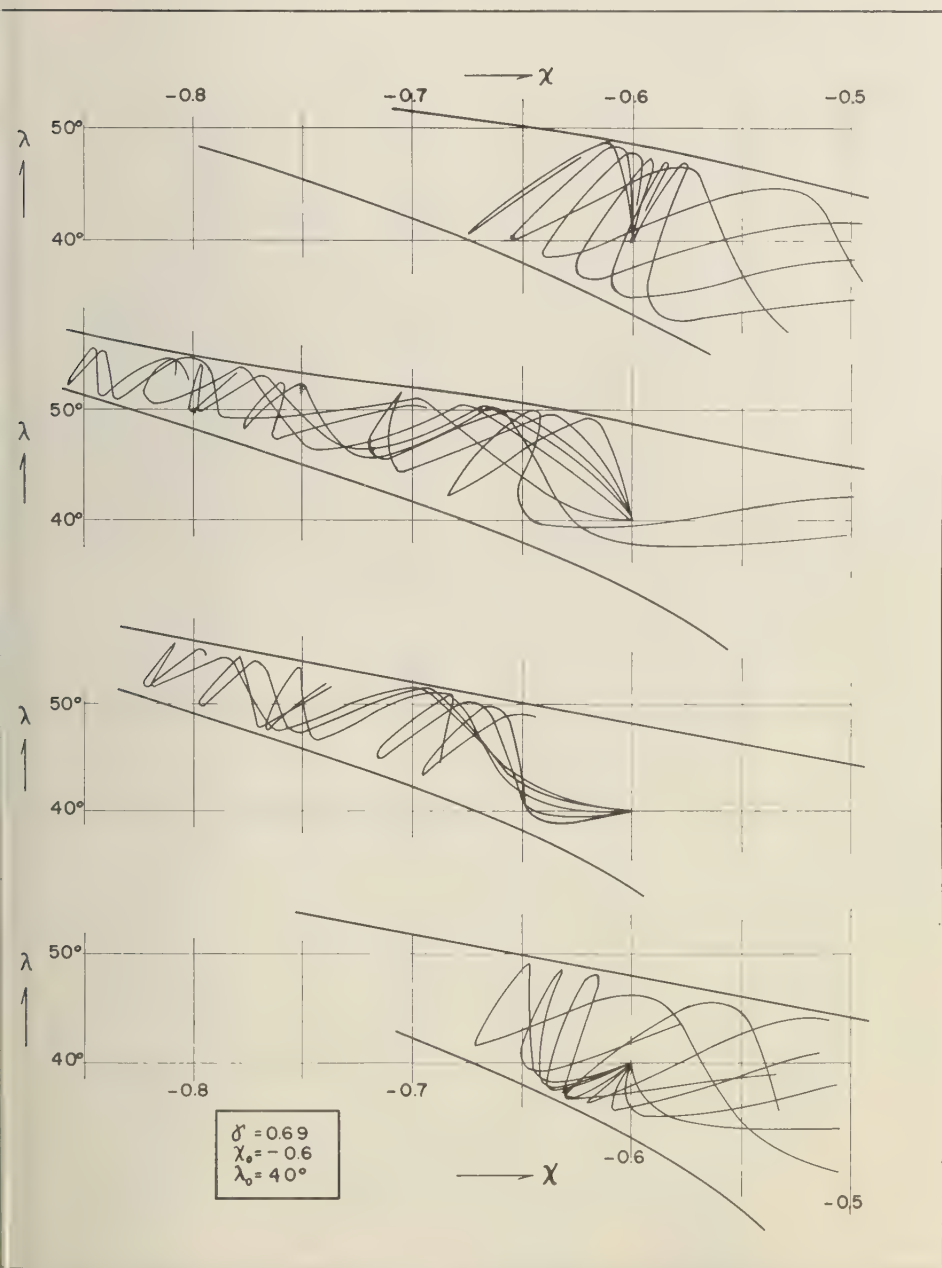


Fig. 1. Some members of pencil of trajectories issuing from the point $x = -0.08$, $\lambda = 10^\circ$, for $\gamma = 0.92$. The pencil is continued in Figure 2.



g. 3. Some members of pencil of trajectories issuing from the point $x = -0.6$, $\lambda = 40^\circ$, for $\gamma = 0.69$.
 (The pencil has been drawn in parts to reduce difficulty in tracing trajectories.)

Now if γ is allowed to vary (r_s , r_e , and λ_s being held constant) an area of darkness in the phase diagram will be generated by the corresponding movement and continuous deformation of the vertical line segment representing the principal shadow region. 'Area of darkness' means that every point in the area corresponds to a physically unallowable direction of arrival. Such areas will also be called principal shadow regions, and the curves bounding them principal shadow cones (the curves being projections of three-dimensional cones with vertices at the point of observation).

For fixed particle rigidity and latitude of observation, the effect of the variation of altitude of observation is especially interesting. It is guaranteed that a principal shadow region for one value of each of r_s and λ_s will begin at $r_e = r_s$ (zero altitude) by being identical with the simple shadow region and will vary continuously from its initial state as the altitude increases.

An extension of the definition. The previously given definition hinges on the requirement that observation points be $P1$ or $P2$. However, as already remarked, for large enough values of r_s and λ_s (or rigidity), or for large enough altitudes even for small rigidities, it cannot be safely assumed that observation points are $P1$ or $P2$. Nevertheless, values of rigidity and altitude of practical interest require that such cases be considered; in order to deal with them, extension must be made of the simple, well-supported definitions given above.

Let values of r_s , λ_s , and γ be selected such that the point x_s , λ_s corresponding to these values is surely $P1$ or $P2$. If the surface of the earth is inserted left of this point at $x = x_s$, there will exist, in general, a principal shadow region for the chosen configuration. If now γ is allowed to increase, or the altitude to increase, both, while r_s and λ_s are held constant, then, at least over some range of increase of those quantities, the continuous changes in them will be accompanied by continuous changes in the principal shadow regions. In particular, if the altitude or γ or both is allowed to continue to increase until the points x_s , λ_s encountered are no longer $P1$ or $P2$, there will yet continue to be generated in continuous fashion a sequence of $T4$ which will continue to bound continua of forbidden trajectories. If any observation point which arises for consideration is not $P1$ or $P2$,

then, there can yet in principle be selected from among the $T4$ emanating from the point those which bound continua of forbidden trajectories such that the forbidden continua are natural extensions of the principal shadow regions which existed at lower altitudes or γ values.

In short, a principal shadow region will now be taken for any point x_s , λ_s and values of r_s and γ to be a region as previously defined, if the point is $P1$ or $P2$; otherwise it will be that continuum of forbidden trajectories which can be traced as the altitude or γ or both are reduced until a continuum of trajectories is reached which is a principal shadow region, as first defined.

THE PENUMBRAL DARK REGIONS

In general, for observation points above the earth the principal shadow regions do not contain all trajectories which intersect the earth, as examination of computed trajectories reveals and, indeed, as intuition suggests. Such forbidden trajectories outside principal shadow regions must generally make continuous transitions to penumbral forbidden trajectories of the kind contemplated in earth-bound geomagnetic theory, as the observation point is allowed to fall to the surface of the earth.

Penumbral dark regions still arise in much the same way as in the older theory; when the earth is inserted in the x , λ plane to the left of the point x_s , λ_s , some trajectories out of that point but not inside principal shadow regions cut the earth, and, because of the continuous deformation of the trajectories with change of the initial conditions, the trajectories affected will lie in bands bounded by grazing trajectories. However, the penumbral forbidden trajectories and the trajectories bounding penumbral regions will now be of somewhat different character than those in the earth-bound theory. Inspection shows that some penumbral trajectories now need not have at least two re-entrant sections, as in the older theory, but rather some need have no re-entrant sections. It may also be recalled that Schremp [1938a, b] has shown that penumbral trajectories can exist only for values of γ between 0.788541 and 1. It now appears that penumbral trajectories can exist for all values of γ to 1, for observation points at non-zero altitudes.

This suggests that the older mode of definition of penumbral regions is best abandoned for

present purposes. Instead of framing a definition resting on specification of characteristics of trajectories (which would be necessary prior to computation of penumbral regions), it seems desirable to refer to any dark region in a cone diagram which is outside a principal shadow region but is not a region of Störmer darkness as a penumbral region.

COMPUTATION OF PRINCIPAL SHADOW CONES

Description of method. An expeditious method for the computation of principal shadow cones is suggested by the method used for the computation of simple shadow cones [Kasper, 1959], since, in the present case too, families of trajectories are needed with points of tangency at the earth which show similarities to the families of trajectories of the second kind used for simple shadow cones. However, the trajectories required now must have absolute rather than relative minima in x at their points of tangency with the earth, and long segments of them are needed so that they may pass through observation points at great altitudes. Worse yet, vastly more trajectories need to be computed than in the relatively simple case of simple shadow cones.

The general method used will be described briefly. Families of trajectories of the fourth kind were first computed by numerical integration of the equations of motion. Each family pertained to one value of each of x_0 and γ . Along each line for which $x = x_0$ values of λ were selected for beginning individual trajectories. In each family some 10 to 20 trajectories starting vertically upward and some 6 to 12 starting vertically downward were computed. These were not uniformly distributed in latitude but were condensed in the region of the inner branch of the curve $P = 0$, it having been found that more useful data for the cones resulted from this arrangement. Only trajectories starting at northern latitudes were computed. Some of these trajectories were found to yield cone data at southern latitudes; such trajectories were used to provide cone data at corresponding northern latitudes by reflecting them in the x, λ plane about the equator. This operation is justified by the symmetry of the x, λ plane.

The values of r_0 used were 0.3, 0.4, 0.6, and 0.8. Values less than 0.3 were not necessary because, in such cases, cone diagrams of interest are either entirely Störmer-dark or nearly so. Values

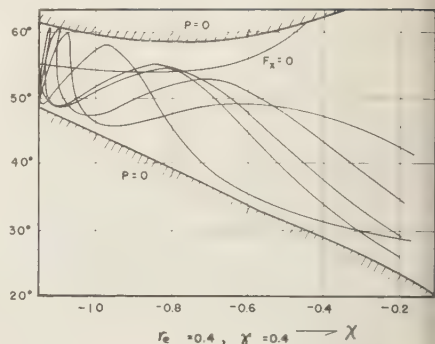


Fig. 5. Some members of a family of trajectories vertically tangent at the earth, for $r_0 = 0.4$, $\gamma = 0.4$.

larger than 0.8 were not used because high rigidity particles 'see' the earth effectively as geometric sphere. The values of γ used were from 0.1, to 1.0 in steps of 0.1.

Most of the trajectories were computed with the IBM 704 computer at the General Motors Technical Center, with a program written by the author. Others were computed with the Martin Aircraft IBM 704 by a RIAS, Inc. group in conjunction with computer personnel of Martin Aircraft Co., Baltimore, Maryland. Figures 5 and 6 show a few members of two of the families computed, and are meant to indicate the general natures of the trajectories dealt with and to illustrate the discussion to follow.

Each trajectory gives, at each crossing of a particular latitude, values of the altitude and

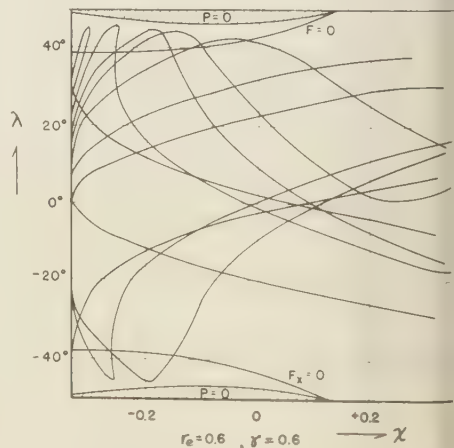


Fig. 6. Same as Figure 5, but for $r_0 = 0.6$, $\gamma = 0.6$.

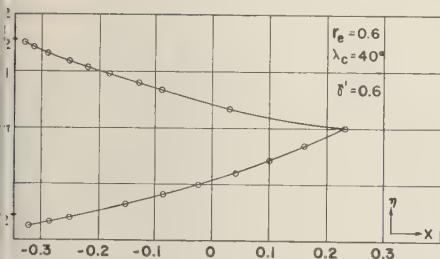


Fig. 7. Example of curve showing the angle η as function of x_s , for $r_s = 0.6$, $\lambda_s = 40^\circ$, $\gamma = 0.6$.

the angle η there, and therefore gives a point on a cone diagram lying on a principal shadow cone, if the trajectory is of a kind which makes up a principal shadow region (a question which was settled by inspection and application of suitable criteria to the trajectories). The altitudes chosen for use were from 0° to 70° north, in steps of 10° .

However, a trajectory started randomly at the earth will cross a given latitude at an altitude which can be discovered only after the trajectory has been computed. A search process for finding each trajectory the right initial conditions to guarantee a crossing at one of a set of preselected altitudes would have been very prohibitive. Therefore, in order to collect cone data at preselected altitudes and preselected latitudes, the trajectories were started at the earth without such a search process, and the data pertaining to the crossings by the trajectories of the preselected latitudes were subjected to interpolations in order to find corresponding data at the altitudes wanted. The x_s , x_e , λ_s , γ , and η data printed out by the computers were grouped so that there were in each group the data for one value of each of γ , x_s , and λ_s . Then the η , x_e data in each group were plotted graphically, and smooth curves were drawn through the plotted points. Figure 7 shows a sample of such an η , x_e curve for one value of each of r_s , γ , and λ_s . The interpolations for values of η at preselected altitudes were then made by reading off the altitudes from such curves. The altitude values chosen, in terms of $(x_e - x_s)$, ranged from 0 to 10 in some 20 increments.

On hand after these operations was a collection of cone diagram points at preselected values of particle rigidity, latitude, and altitude. However, these data were in the form of sets of values

of x_e , x_s , λ_s , γ , and η . Conversion into cone diagram coordinates $\sin \theta$, $\cos \theta \sin \eta$ was done by hand substitution into the formulas relating the coordinates to the quantities available; this was not a small job. Finally, the computed points were plotted in cone diagrams, and smooth curves were drawn through the points.

Discussion of the cones. The computed cones were in the hundreds, and only a sampling will be given here. These are presented in Figures 8 through 11; representative rigidities, latitudes, and altitudes have been selected. These figures are projections of the observer's sphere of directions onto the observer's horizontal plane. The origin is at the center in each case, and the horizontal and vertical axes are, respectively, the $\sin \theta$ and $\sin \eta \cos \theta$ axes; grid lines for the values 0, 0.2, 0.4, 0.6, and 0.8 have been included but not labeled (to avoid cluttering the figures) for the positive and negative directions for both coordinates. Curves shown in the diagrams are computed principal shadow cones. Where Störmer cutoffs are important some of these have been included for guidance; such cutoffs are shown as heavy vertical lines. These Störmer cutoffs (more properly called cones) are labeled with values of the altitude expressed as numerical values of $(x_e - x_s)$; for an explanation of a simple manner of converting these values into values of the altitude in ordinary units, see the discussion of the corresponding question for the principal shadow cones given below. It is worth remarking that for low particle rigidities the Störmer darkness is very important indeed but the ease with which the effect can be computed has made it seem undesirable to include more than a few Störmer cones for some of the diagrams.

Interpretation of the principal shadow cones will now be discussed. Complete identification of the values of the parameters pertaining to any cone is to be made by referring to the values of r_s and λ_s attached to the figure containing the cone, and to the number attached to the cone itself. The indicated value of r_s is the particle rigidity in Störmers, and λ_s is the geomagnetic latitude of observation; all latitudes shown are northern latitudes. The number labeling each cone is the altitude of observation, given as a value of $(x_e - x_s)$. When altitude is specified in this mode it is unnecessary to settle on a specific (and necessarily moot) value for the

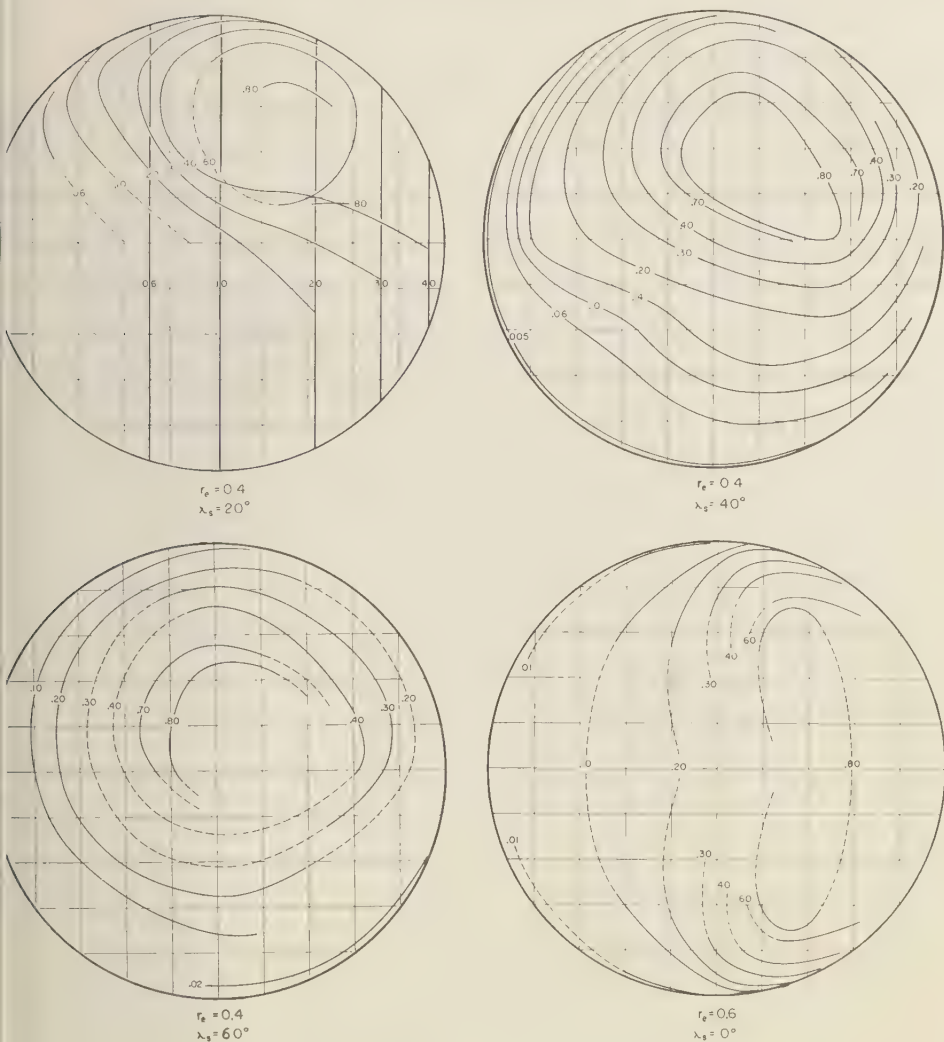


Fig. 9. Selected principal shadow cones.

and has not been shown except in the case of the upper-hemispheric parts of the shadow cones (which are given elsewhere [Kasper, 1959]), and to give them would greatly obfuscate the drawings. Finally, some arcs of principal shadow cones are drawn in fully inked curves and some in dashed curves. The former indicate arcs relatively well fixed by a large number of computed points; the latter indicate arcs less well fixed.

It is useful to be able to make a comparison of any cone with the geometrical shadow of the earth at the same observation point. This can be done easily because the geometrical shadow in a cone diagram is a circle with center at the center of the diagram and with radius ρ_e/ρ_s , which in turn is given very simply by $\exp(x_e - x_s)$, as is easily shown.

The characteristic shift of the cones to the north and east (as compared with the corresponding geometrical shadows) is easily under-

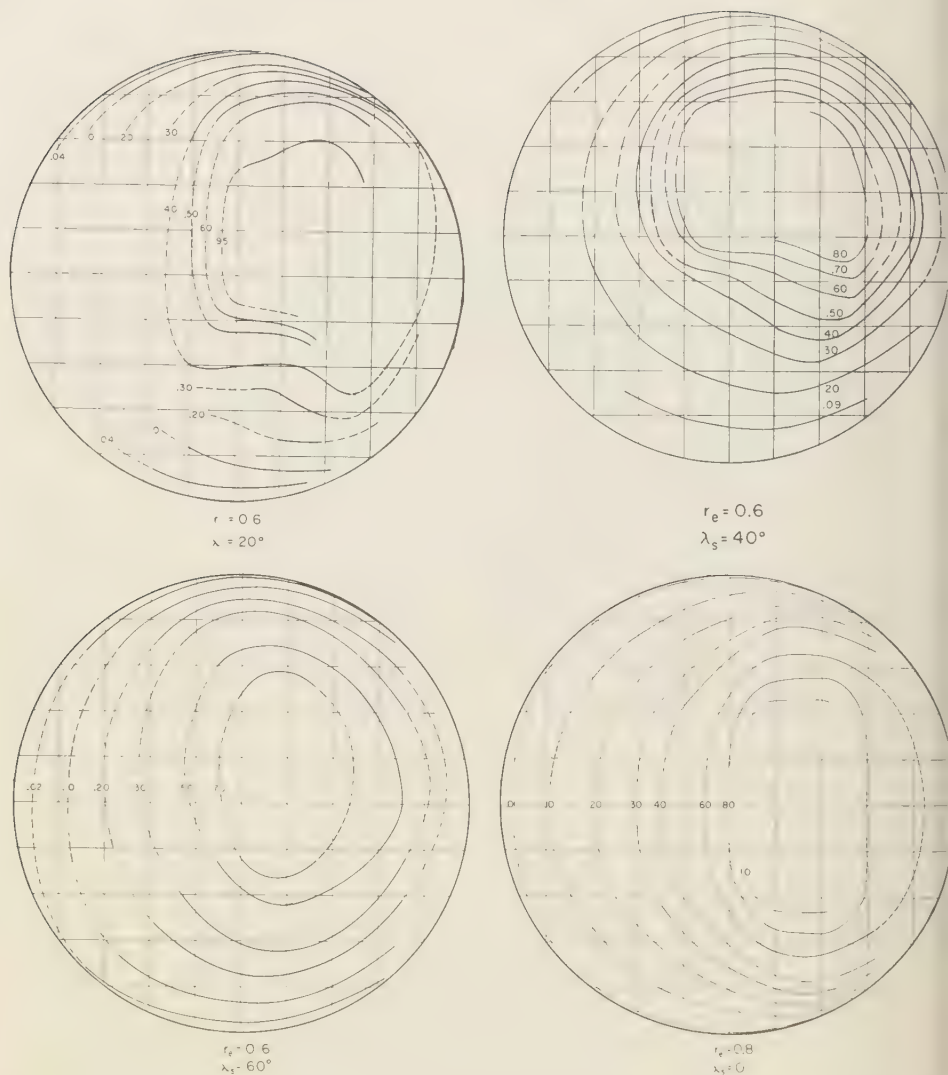


Fig. 10. Selected principal shadow cones.

stood as being due to the spatial forms of the magnetic lines of force of the earth's field and the direction of deflection of positively charged particles as given by the Lorentz force. The symmetry of the equatorial cones with respect to the $\sin \theta$ axes of the diagrams corresponds to the symmetry of the x, λ plane with respect to the equator. A loss of such symmetry with increasing latitude and a subsequent tendency

with further increase of latitude for the cones to move toward the centers and to become circular reflects their approach to the kind of symmetry appropriate for the polar regions. The lobe shown by many cones at low latitudes are to be ascribed to the barrier formed in the x, λ plane by the inner branch of the curve $P = 0$, which forces trajectories out of low latitude points to deflect upward and downward.

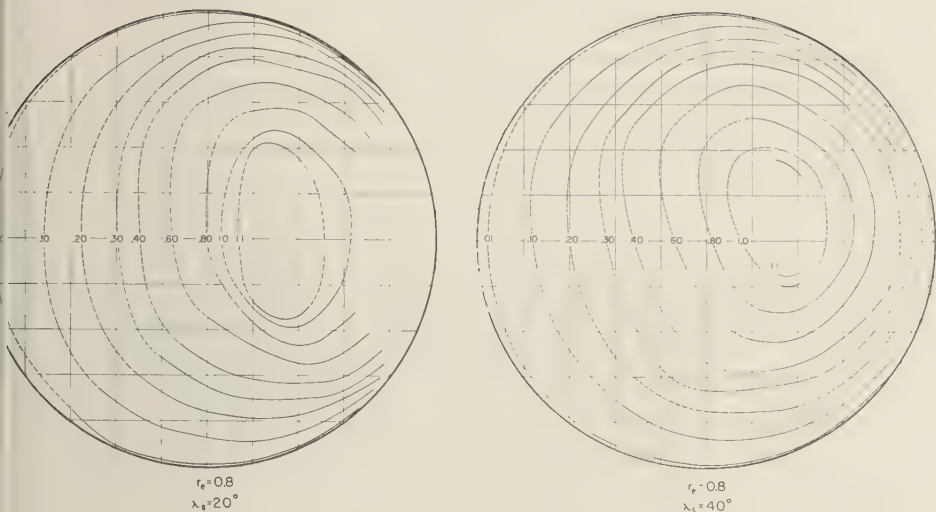


Fig. 11. Selected principal shadow cones.

At low altitudes and high rigidities there is agreement between the sizes of the cones and the sizes of the corresponding geometrical shadows, as is to be expected. At high altitudes there is a tendency for principal shadow cones to become smaller than the geometrical shadow; this speaks for a deflection of particles into the observer's sky, an effect noticeable on examination of trajectories.

With fixed altitude and latitude a decrease in the value of r_g (rigidity) is to be expected to be accompanied by an increase in the total amount of darkness in the observer's sky. However, such a decrease in r_g need not result in an increase in the size of the principal shadow cone. There is, in fact, sometimes a concomitant decrease in the shadow cone, but in such cases a suitable overcompensating increase in the amount of umbral darkness may accompany the decrease in r_g , so that the total amount of darkness increases.

It is planned to perform integrations over computed principal shadow cones and Störmer cones in order to exhibit the fraction of the sky considered dark by them, as this depends on altitude for each rigidity and latitude.

Acknowledgments. The author wishes to thank A. Van Allen for suggesting the value of this undertaking and for much material aid, and E. C. Wy for long-continued help and suggestions. He is grateful to the General Motors Corporation,

RIAS, Inc., and Martin Aircraft Co. for access to their computer facilities, to R. C. Herman of the General Motors Research Staff for material aid, to the Aerojet General Corporation for a grant covering part of the time spent on this project, and to the numerous persons of Iowa City, Iowa, who carried out tedious hand numerical work. This project was assisted by the National Science Foundation, IGY Project 32.1.

REFERENCES

- Hutner, R. A., On the penumbra of cosmic radiation, *Phys. Rev.*, **55**, 15-23, 1939.
- Hutner, R. A., The penumbra at geomagnetic latitude 20° and the energy spectrum of primary cosmic radiation, *Phys. Rev.*, **55**, 614-620, 1939.
- Kasper, J. E., The earth's simple shadow effect on cosmic radiation, *Suppl. Nuovo cimento*, **11**, 1-26, 1959.
- Lemaitre, G., and M. S. Vallarta, Geomagnetic analysis of cosmic radiation, *Phys. Rev.*, **49**, 719-726, 1936.
- Lemaitre, G., and M. S. Vallarta, On the allowed cone of cosmic radiation, *Phys. Rev.*, **50**, 493-504, 1936.
- Schremp, E. J., I. General theory of the earth's shadow effect on cosmic radiation. II. The simple shadow cone of cosmic radiation, *Phys. Rev.*, **54**, 153-162, 1938.
- Störmer, C., *The Polar Aurora*, Clarendon Press, Oxford, 1955.
- Vallarta, M. S., *Univ. Toronto Studies, Appl. Math. Ser.*, no. 3, University of Toronto Press, 1938.

(Manuscript received July 30, 1958; revised October 22, 1959.)

A Radial Rocket Survey of the Distant Geomagnetic Field¹

C. P. SONETT, D. L. JUDGE, A. R. SIMS, AND J. M. KELSO

*Space Technology Laboratories, Inc.
Los Angeles 45, California*

Abstract. A survey magnetometer carried aboard the Air Force Pioneer I space probe obtained data over the interval 3.7 to 7 and 12.3 to 14.8 geocentric radii. The day was unusually quiet magnetically ($A_p \sim 5$) with a long quiet prior history. The data indicate an inverse-cube-field decrease in the region of 3.7 to 13.6 radii, where termination takes place, with a subsequent decrease to 5×10^{-8} gauss. This value is examined in the light of possible interplanetary gas activity. The surprisingly distant geomagnetic cutoff suggests a very low gas pressure on the day of the flight. Increasingly large (fractionally) fluctuations were observed with increasing radii. A gross variability in the vestigial field would suggest hydro-magnetic activity or a complex gas cloud structure.

Introduction. A search-coil magnetometer was included among the instrumentation of the Air Force Pioneer I probe, launched October 11, 1958, to survey the magnetic state of the medium between earth and moon, as well as to test for interplanetary magnetic field. Although the objective of a lunar orbit was not achieved, a considerable quantity of data was obtained over two regions of the ascending leg of the orbit. These were from 3.7 to 7 geocentric radii, and from 12.3 to 14.8 radii, measured from the center of the earth. The body of this paper is concerned with these data and their interpretation in the light of the many system parameters involved. The two regions of data will henceforth be designated as A and B, respectively.

These data provide an opportunity to examine one of the numerous hypotheses that have appeared in the past few years concerning the physical state of the outer atmosphere. Perhaps the most prevalent has been the concept of a solar wind transporting both magnetic energy and corpuscular radiation outward from the sun at possibly 2.5 astronomical units [Parker, 1958a]. Evidence for such a streaming wind has been obtained by Biermann [1957] from examination of the drag forces acting upon a certain class of comet tail. His data imply a particle density of 10^4 ion pairs/cm³ composing a Lorentz gas of fully ionized hydrogen). (This is for days

where agitation of the surface field is noted only in the auroral zones, i.e., 'geomagnetically quiet.') Kiepenheuer [1953], on the other hand, estimates a density of 0.5 to 6.0 cm⁻³ and related velocities of 300 to 1000 km/sec. The photometry of the zodiacal light by Behr and Siedentopf [1953] indicated an electron density of 600 cm⁻³ at 1 astronomical unit in approximate agreement with estimates of the evaporation loss of the corona [van de Hulst, 1950; Kiepenheuer, 1951].

A different approach was taken by Hoyle [1956], who has assumed that the orbital motion of the earth results in an interaction with a stationary plasma. Nevertheless, as he points out, such an interaction should have a complex involvement with the field of the earth. The arguments involving an interaction between the geomagnetic field and an interplanetary plasma have led to a number of further suggestions regarding the detailed behavior of the boundary [Parker, 1958b] and the proposed subsequent generation of hydromagnetic waves [Dessler, 1958]. We shall consider these processes in the following sections. It is appropriate to point out here that there is little correlation, to date, between these observations and the state of solar activity as reflected either by direct observations or by possible secondary effects near the earth's surface.

Constant agitation of the surface geomagnetic field is a well established phenomenon and has been studied extensively. Dungey [1958] has suggested that this effect is caused by waves generated at a plasma interface. Dessler thinks

¹This program was carried out under the auspices of the National Aeronautics and Space Administration.

that the amplitude above the ionosphere may be some 10^2 greater than the expected surface value, because of dissipative losses and reflection in the ionosphere.

The interplanetary magnetic field may play an important role in the propagation of plasma material from the sun and is probably involved in the propagation of cosmic rays. On average days, cosmic-ray data indicate an upper bound of some 3×10^{-4} gauss. If we utilize the solar-field transport of *Parker* [1958a] the solar field at 200 R should be $2.5 \times 10^{-5} B_0$, where B_0 is the effective solar field. (B , the magnetic induction vector, is used throughout this paper since the units have the same numerical value as H in emu.) *Alfvén* [1958] also assumes an inverse-square decrease of the solar field in the absence of flares and other similar events. *Morrison* [1956] has postulated that a turbulent structure is associated with the over-all field, and a generalized heliocentric turbulent halo has been invoked by *Parker* [1958a] to explain certain time delays associated with the onset of flares and the reception of related corpuscular radiation on earth. Clearly, we cannot hope to be able to separate simply ambient and sporadic properties, and it may well turn out that this will become a basic problem in experiments involving electrodynamics.

Description of the magnetometer. The magnetometer used in this experiment consisted of a coil of 30,000 turns of No. 40 copper wire, wound on a nickel-iron alloy core. The length-to-diameter ratio of the core was approximately 40-1, and the coil extended only over the center 2 inches of the 10-inch-long core material. The emf generated was applied to an amplifier having a center frequency of 2.0 cps and a passband of 2.0 cps [*Judge and others*, 1959]. The coil was fixed in the vehicle with its plane collinear with the spin axis, and, therefore, only the component of B lying in the plane of the vehicle's equator was measured. This component will henceforth be designated as B_1 . Since this type of magnetometer measures the time rate of change of a component of the magnetic intensity vector, errors are introduced into the signal if B is time variant with a frequency of the order of the spin rate. Such regions were found only in the data for region B.

We shall briefly describe the signal emf generated by the search coil [*Judge and others*,

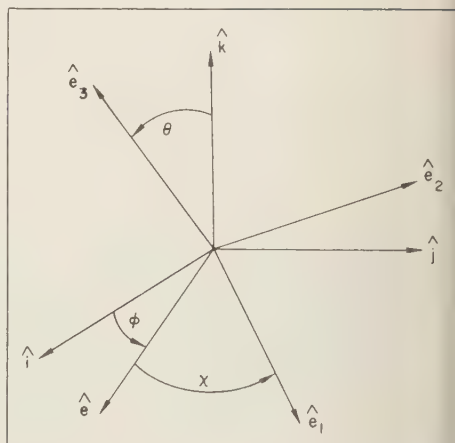


Fig. 1. Coordinate reference triads for the magnetometer geometry (generalized to include precession). The triad $\hat{i}\hat{j}\hat{k}$ is body-centered and inertial; the triad $\hat{e}_1\hat{e}_2\hat{e}_3$ rotates with the vehicle. The axis of angular momentum lies along \hat{k} ; the axis of spin along \hat{e}_3 . \hat{e} is a unit vector along the line of node. The plane of the magnetometer coil has its normal directed along \hat{e}_1 . For the case considered in this paper, \hat{e}_2 and \hat{k} are collinear, i.e., condition of no precession.

1959; *Sims*, 1959]. Define two coordinate systems, both right-handed Cartesian and centered on the vehicle as shown in Figure 1. The set $(\hat{i}\hat{j}\hat{k})$ is inertial and centered in the vehicle, and the set $(\hat{e}_1\hat{e}_2\hat{e}_3)$ is fixed in the vehicle. For a nonprecessing vehicle the axes \hat{k} and \hat{e}_3 are collinear and in the direction of the vehicle angular momentum. Thus, the emf generated by vehicle motion is given by

$$E = -NA(d/dt)(\mathbf{B} \cdot \hat{e}_1) \quad (1)$$

where $\mathbf{B} = B_x\hat{i} + B_y\hat{j} + B_z\hat{k}$, the field described in the inertial set, and NA is the product of the number of coil turns and the coil area. For a nonprecessing vehicle and $\dot{\mathbf{B}} = 0$,

$$E = -NA\sqrt{B_x^2 + B_y^2}\dot{\chi}\sin(\chi - \epsilon) \quad (2)$$

where χ is the third Euler angle (using Goldstein's notation [Goldstein, 1951]) and ϵ an arbitrary phase. For a precessing vehicle,

$$E = -NA\left[\frac{\omega_3}{2}\sqrt{B_x^2 + B_y^2} + \left(\frac{I_3}{I_1}\sin\theta_0\tan\theta_0\right)\right] \quad (3)$$

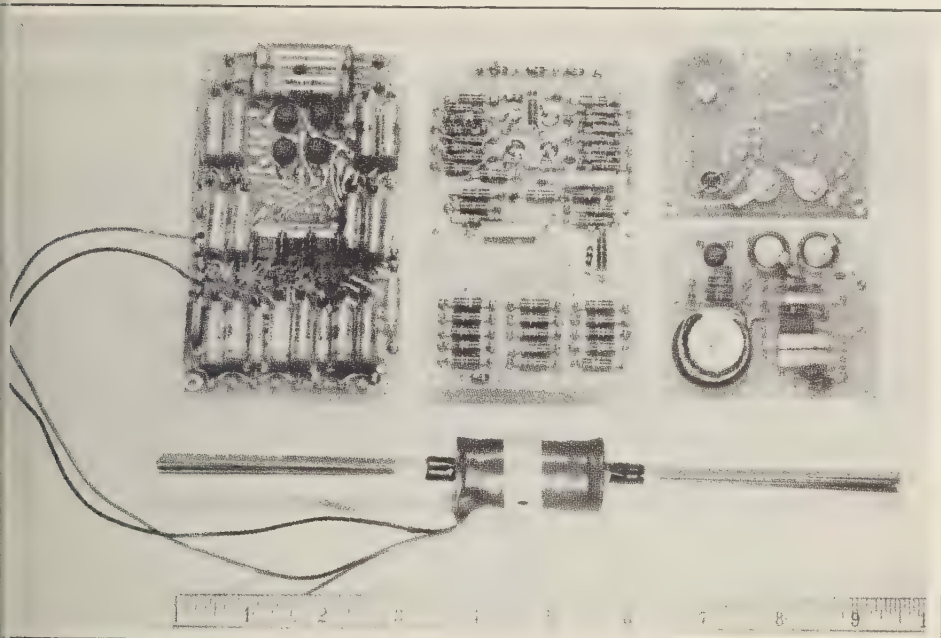


Fig. 2. Magnetometer system. In the upper left-hand corner is a front view of the magnetometer amplifier. The back view of the magnetometer amplifier is shown in the center. In the upper right-hand corner a front and a back view of the subcarrier oscillator, and below is shown the search coil.

$$\begin{aligned}
 & + \cos \theta_0 + 1 \Big) \sin (\phi + \chi - \epsilon) \\
 & + \frac{\omega_3}{2} \sqrt{B_x^2 + B_y^2} \\
 & \cdot \left(\frac{I_3}{I_1} \sin \theta_0 \tan \theta_0 \right. \\
 & \left. + \cos \theta_0 - 1 \right) \sin (\phi - \chi - \epsilon) \\
 & - B_x \dot{\chi} \sin \theta \cos \chi \Big] \quad (3)
 \end{aligned}$$

analysis of the data, including a spectrum of spin-modulated RF carrier, showed the precession-induced frequencies $(\dot{\theta} + \dot{\chi})$ and $(\dot{\theta} - \dot{\chi})$ to be absent. This, together with the absence of a precessional damper aboard the vehicle, is considered adequate evidence of angular momentum about only the spin axis.

The magnetometer system is shown in Figure 1. The output of the amplifier, a sinusoid varying at the characteristic spin-rate frequency, was used to drive a subcarrier oscillator, the subcarrier output being given by the usual frequency

modulation expression

$$A \sin [\omega_c t + (f_d/\omega_m) \sin \omega_m t] \quad (4)$$

where ω_c is the carrier angular frequency, ω_m the spin angular frequency, and f_d the maximum swing of the subcarrier frequency. This subcarrier constituted one channel of the total telemetry system and was combined in a multiplexing amplifier with the subcarrier of other experiments [Sonett and others, 1959a]. The output was used to phase-modulate a 350-mw transmitter operating at 108.06 Mc/s. Therefore, the envelope of the magnetometer tone provided a measure of the value of B_1 . The phase of the signal indicated changes in orientation of B_1 to within a cone about the spin axis, \hat{k} .

Calibration procedure. The laboratory calibration procedure consisted of determining the coil constant, $K = NA$, in terms of a standard coil. The standard consisted of a large air-core solenoid whose turns times area could be accurately calculated. The earth's field was utilized, and a carefully aligned jig was used to spin the two coils, both mounted on the jig, at

the expected vehicle rate so as to generate an emf. The effective magnetic field could be varied by placing the spin axis at an angle to the geomagnetic field. Checks were made to ensure a tolerable magnetic gradient in the region of the jig so as to provide a sufficiently uniform field for the calibration. Since the search-coil constant is given by $(E_c/E_{sc})K_{sc}$, where E_c/E_{sc} is the peak-voltage ratio for the two coils and K_{sc} the standard coil constant, we note that aside from nonlinearity due to a functional dependence of the form $\mu = \mu(H)$, or $\Delta|B| \neq 0$, the measurement is exact. The relation $d\mu/dH \cong 0$ was verified for small fields as discussed below, and extensive environmental tests of the search coil showed no history-dependence or variation of μ with field.

The amplifier was calibrated separately by driving it from a source which duplicated the impedance of the coil. Gain curves were made over a wide range of temperature and over the passband of the amplifier. The final calibration of the complete magnetometer was done by exciting the equipment in the payload and determining the output throughout the submodulation so as to include the subcarrier modulator and multiplexing amplifier characteristics. A typical calibration curve is shown in Figure 3.

Since the magnetometer was only a-c sensitive and had a limited passband, currents within the vehicle were not sensed. Little ferromagnetic material was used in the vehicle. What there was would have influenced field measurements, by a change in the magnetic induction; permanent dipoles were unsensed by the magnetometer. Such changes were investigated and found to be unobservable for the amount and geometry of such material.

The effects of variable and small fields upon the permeability of the search-coil core were carefully considered. At most, the change in permeability, as reflected in harmonic generation in the output emf, was less than 2 per cent. A theoretical analysis of such effects was also made [Margolis, 1958]. Since the core had a large length-to-cross-sectional geometry, it can be represented by a prolate spheroid of extreme eccentricity, where [Stratton, 1941]

$$B_x = \frac{\mu_0 \mu H_{0x}}{1 - \frac{bc}{2a^2 e} \left(2e + \ln \frac{1+e}{1-e} \right)} \quad (5)$$

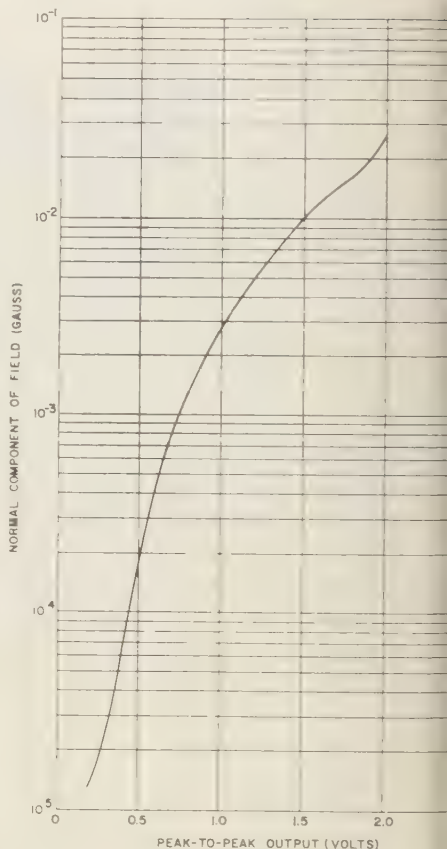


Fig. 3. Field strength versus amplifier output.

where a is the long axis, b and c the transverse axes and $a \gg b = c$, B_x is the internal μ -modified field in the long direction, e the eccentricity, and H_{0x} the magnetic intensity in the x direction. Thus,

$$dB_x = H_x d\mu_{eff} = H_x d\mu/40 \quad (6)$$

showing that, for $\mu = 2 \times 10^4$, an effective reduction of $d\mu$ by a factor of 40 occurs. Therefore, real changes in permeability would reflect into B_1 reduced by a factor of 40.

The input circuit of the magnetometer amplifier was d-c coupled to the search coil. A direct current of approximately 0.1 milliampere biased the core to 1 gauss. Thus, the incremental hysteresis loop was offset. This in conjunction with the high degree of linearity of the core for the fields present assured that the origin

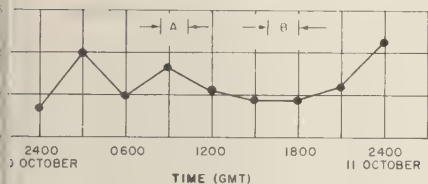


Fig. 4a. Magnetic planetary K index for the time interval about the flight data. The times of regions A and B are shown.

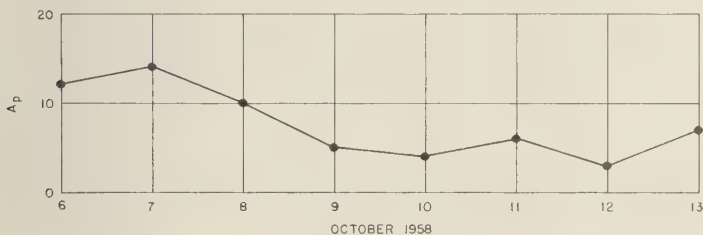


Fig. 4b. The planetary A index from October 6 to 13, 1958. Data taken from Part B, Solar-geophysical Data, issued December 1958, U. S. Department of Commerce, National Bureau of Standards, Central Propagation Laboratory, Boulder, Colorado.

B - H diagram was always avoided for small B , and so the question of operation of magnetic material in the region of zero magnetization was not present.

Trajectory and other system parameters. The flight time and trajectory of the vehicle were determined by orbital and powered-flight constraints of a complex nature requiring optimization of a number of rocket system parameters. Because of these primary requirements, a geological basis on which to establish a firing time was not practicable. As the data have since demonstrated, the extremely quiet nature of the day was a fortuitous occurrence, the untangling of electromagnetic data of this type being a complex problem even in such a case.

The Pioneer I vehicle was launched at 0342 UT on October 11, 1958. The planetary A index had been steadily decreasing since the beginning of the month, and on the day of the launch had dropped to a very low value. (See Figs. 4a and b.) The trajectory was designed so that appropriate intersection with the lunar orbit point would occur. This required that the vehicle be fired into a specified plane lying between the ecliptic and that containing the

earth-moon orbit. The actual trajectory flown by the vehicle was determined from tracking information supplied by stations at Cape Canaveral, Florida; Jodrell Bank, Great Britain; Millstone Hill, Massachusetts; and the STL tracking site on the island of Hawaii. Information transmitted to Los Angeles by teletype was supplied to an IBM 704 computer, which, in turn, computed the trajectory and the vehicle orientation in an earth-fixed coordinate system. In addition, the expected value of B_{\perp} was computed by basing it upon an eccentric dipole



Fig. 5. Pioneer-type payload showing the micro-meteorite diaphragm and one leg of the dipole antenna. The magnetometer search coil is mounted circumferentially and cannot be seen from this view.



Fig. 6. The angle, ξ , defined by the direction of the expected magnetic field and the axis of angular momentum and spin of the vehicle as a function of the geocentric radius.

model with the first eight Gauss coefficients [Vestine, 1953] folded into the trajectory program.

The vehicle consisted of a plastic laminate shell of an approximately toroidal shape, as shown in Figure 5. For reasons connected with the guidance system, the vehicle was spun at an angular rate of 1.835 rps. Since the vehicle had intrinsic spin, it was particularly appropriate to use a search-coil magnetometer; it was decided to make use only of the intrinsic spin of the vehicle, because of serious problems associated with the coupling of angular momentum of a spinning vehicle and a magnetometer turning in the vehicle frame. Thus, the magnetometer coil was fixed in the rotating frame of the vehicle. The spin rate of the vehicle was constant throughout the data interval to within 0.005 cps. The antenna pattern was sufficiently anisotropic so as to provide a noticeable spin-rate modulation of the RF carrier. This was a useful phase reference, then, for testing the data for phase variation in the azimuthal direction, i.e., in the plane $\hat{r}\hat{\theta}$.

The orientation of the vehicle axis was, fortuitously, almost perpendicular to the extrapolated field, B , over most of the data intervals.

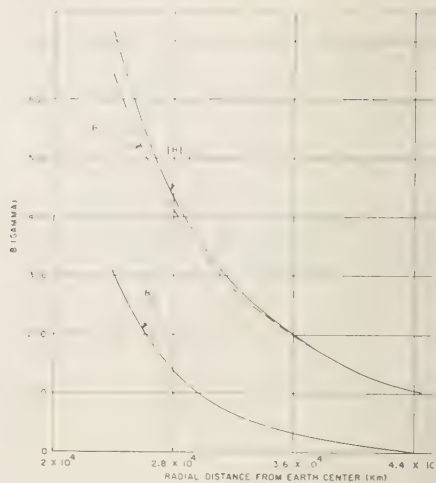


Fig. 7. The magnitude of the magnetic induction $|B|$, and the components B_1 and B_2 as a function of radial distance from earth center.

The angle, ξ , between the vehicle axis and extrapolated field for region A is shown in Figure 6 as a function of radial distance from the center of the earth. The convergence to a value of $\pi/2$ is clearly visible. This is also demonstrated in Figure 7, where the field components and total field are shown for region A. Thus B_1 is a good measure of $|B|$ over most of the trajectory.

From lift-off to 3.7 geocentric radii the Manchester antenna tracked on a side lobe because of a deviation of the trajectory from its nominal value. Thus, the signal-to-noise ratio was inadequate to demodulate the magnetometer signal during this interval. Sophisticated demodulation procedures are still being investigated to attempt to obtain usable information during this interval. At 3.7 geocentric radii the antenna angle was corrected and main lobe tracking commenced with good signal. From 7 to 10 geocentric radii a subcarrier oscillator interaction occurred with loss of magnetometer signal. This was determined from examination of the subcarrier spectrum by means of an antenna spectrum analyzer. The signal returned at 10 radii, and from this time until the end of tracking day at Manchester signals were obtained continuously except during intervals of antenna slewing (used to obtain angular orbit data) and during Doppler breaks. For the periods of Doppler breaks telemetry transmitter was being u

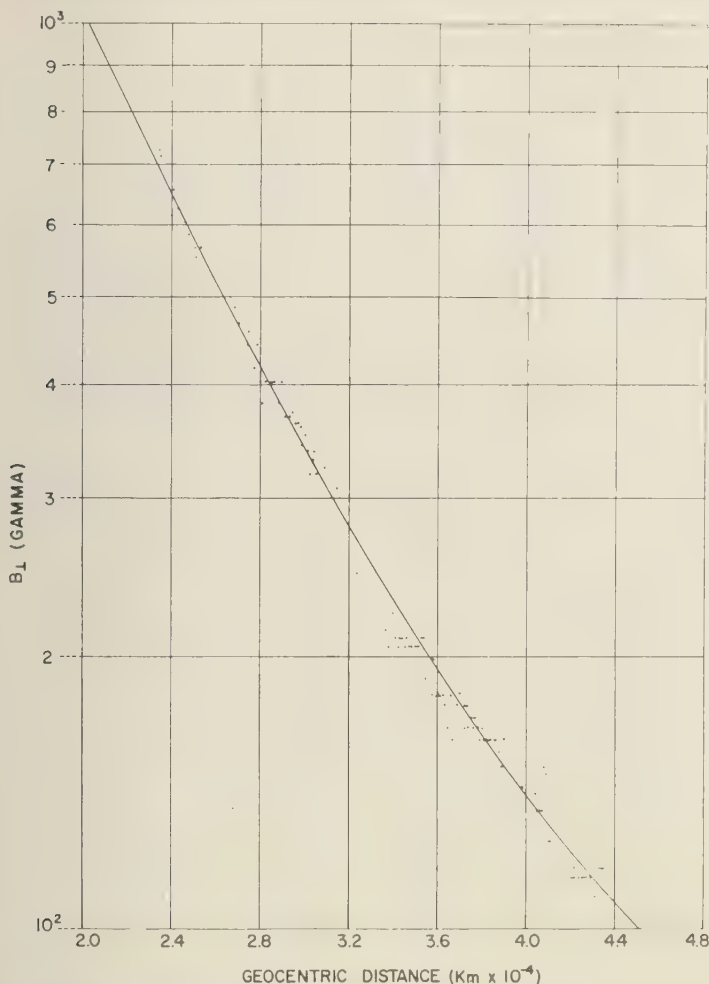


Fig. 8. B_{\perp} as a function of the geocentric distance for region A. The solid line is the expected value B_{\perp} based upon an eccentric dipole and orbital parameters. The validity of the fluctuations is discussed in the text.

the return link of a coherent Doppler velocimeter.

The geomagnetic field for region A. Data for an interval from 2.36 to 4.38×10^4 km were plotted in Figure 8. The agreement with the extrapolated value of B_{\perp} is generally within 1 per cent. As a result of the nonlinear amplifier, a figure of 10 per cent represents a change in frequency excursion of the demodulated subcarrier signal of about 0.5 per cent. The frequency stability of the whole system is the rms composite primarily of (1) the subcarrier oscillator,

(2) magnetic-tape transport mechanisms, and (3) discriminator. The primary contributions are due to the tape recording and playback, these having a combined probable error of ± 1 per cent. The subcarrier-oscillator behavior in the vehicle cannot be predicted. However, the deviation sensitivity rather than center-frequency stability is the parameter of interest. The former would be expected to be stable to within ± 0.1 per cent. The demodulator and the recording equipment were carefully checked by repeated data runs. The validity of the fluctuations shown in detail

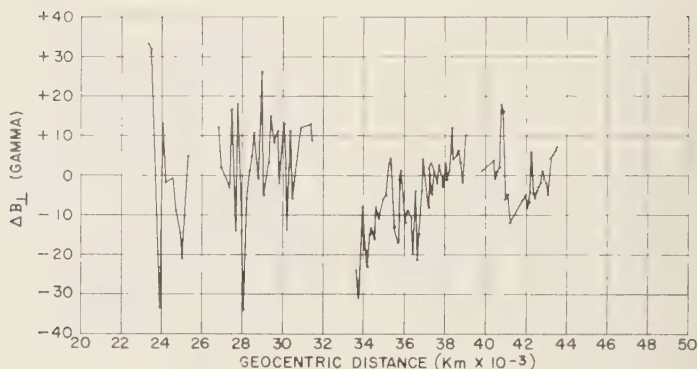


Fig. 9. The differences between the expected and observed values of B_L for region A. Error flags are shown. Errors are discussed in the text.

in Figure 9 cannot be established with confidence, but repeated runs of the tape recordings have established that the gross structure shown in Figure 9 is either real or a result of small drifts in the subcarrier deviation sensitivity. There is a suggestion of a correlation between these fluctuations and those of the ion chamber (Fig. 10) [Rosen, 1959] carried on the same flight. Conclusions about possible fine structure in region A must await a further flight, preferably with a linear magnetometer.

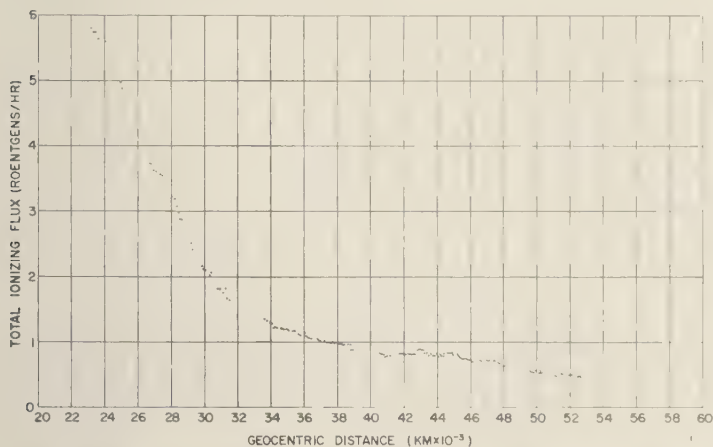
The absolute calibration of the instrument was compromised in the field by telemetry modulation adjustment not under the control of the experimenters. As a result, the validity of the data of Figure 8 must be examined carefully. Since the calibration curve can be represented by a slowly varying function times an exponential (Fig. 3), a scale-factor change of peak-to-peak voltage both translates the final data and changes the power-law dependence of the data. Thus, a unique calibration scale-factor correction translates the data and also provides an inverse-cube fit. It cannot be said with certainty that the data are correct, but it is unlikely that the multiplicative correction applied to the calibration curve (corresponding to the type of change made in the field) would alone change the data as it has. A further comment is that the data for region B, discussed later, must have the same correction. This correction yields the twofold compression noted in region B.

The close general agreement with an extrapolated inverse-cube law must be examined in relation to hypotheses of ambient solar winds.

The data for region A were taken from approximately 45 to 110 minutes after the launch. The vehicle's longitude is shown in Figure 11 to the prime meridian. Since the local time about dawn, the vehicle was under the stagna point for interaction of the earth with a stationary interplanetary plasma. If Hoyle's assumptions are accepted, a magnetostatic compression should have been observed with a twofold increase of the field at the internal side of the boundary. If this supposed boundary did exist, it must have been sufficiently farther out than 4.38×10^4 km so that the field compression (presumably a maximum under the stagna point) was unnoticeable in region A. In relation to the relation of a compression to the calibration, a twofold increase of the field, as discussed above, should drop off in the direction toward the earth. The rate of decrease depends on the extent and geometry of the surface currents that are required to flow in the boundary. Such a dropoff, regardless of its rate, cannot be made compatible with the data because of the apparently logarithmic nature of the calibration of the magnetometer amplifier.

The geomagnetic field for region B (12.5 to 14.8 radii). The signal received during the passage of the vehicle through this region appeared different from that for region A. The generally disturbed character was noted by large-scale changes in magnitude and several directional variations, which will be discussed in a later section.

Because of the disturbances, the data were reduced by means of an IBM 704 computer.



10. The total radiation level as measured by the ion chamber carried in Pioneer I. The data shown cover only region A.

emf peaks and the half-periods of both magnetometer signal and RF signal strengths were transferred to cards for the complete interval of data. Long-time constant AGC effects in the magnetometer amplifier were removed by a special program. Time averages of the corrected field are shown in Figure 12, where each point represents from 80 to 120 data points. As discussed previously, such effects as short-time carrier oscillator and demodulation instability, as well as magnetic-tape frequency errors, must be considered in evaluating the variations in the data. The error flags on the data points represent estimates of error including the effects. It is to be noted with some interest that the data appear to fit the theoretical values quite well, dropping off only in the region of 10^4 km. Plotted in Figure 13 are estimates of geomagnetic cutoff based upon previous data

as mentioned earlier. These figures are to be compared with the data of this experiment, recalling that the day of this flight occurred with a planetary A index of about 5. There is considerable latitude as to the data on the interplanetary gas. Strictly speaking, the cutoff should be considered as the radius at which an interaction takes place. It is more likely that the concept of a cutoff is valid, though a complex mixing of the geomagnetic field and the interplanetary gas may confuse the situation. The term cutoff is used with the above understanding. In examining Figure 13, the estimates of N (particle density) can be fitted to a variable geomagnetic cutoff radius and relative earth-gas velocity. A fixed cutoff (r/r_0) corresponds to a pressure isobar (i.e., horizontal line).

The position of the vehicle was approximately local noon while traversing region B. Thus, it was near the earth-sun line during this time. Interaction with a stationary plasma might require that shear effects be important since, during this time, the vehicle was approximately 90° away from the stagnation point. The effect of shear might well be noticed as an apparent magnetostatic compression, but the actual termination of the field would require a somewhat more arbitrary explanation, perhaps of the type postulated by Parker [1958b] where it must be assumed that, beyond a certain radius, the field contiguity must disappear. On the other hand, the model of a streaming wind would

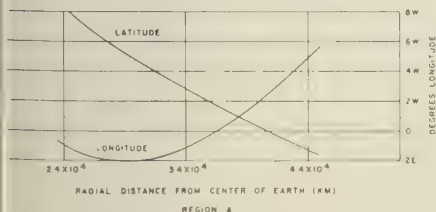


Fig. 11. The geocentric latitude and longitude of vehicle during its traversal of region A. Note that the longitude during this interval was centered at the prime meridian.

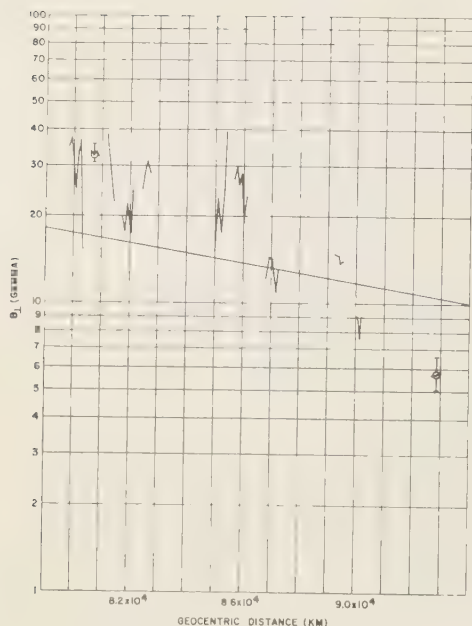


Fig. 12. The magnitude of B_{\perp} for region B (12.3 to 14.6 geocentric radii). The solid curve shows the theoretically expected value of the perpendicular component of B . The two error flags show the variation in error with geocentric distance. See text for assignment of these errors.

require that the shear effects would have occurred near dawn, although unnoticed in our case. Then the data for region B would represent a true compressive effect. It is impossible at present to decide between these alternatives.

The estimated particle density is $1.1 \times 10^2 \text{ cm}^{-3}$, assuming protons and a relative velocity of $3 \times 10^6 \text{ cm/sec}$. This is the case for a stationary interplanetary plasma interacting with the orbiting earth. Clearly, higher estimates of velocity based upon a streaming gas would modify downward the estimated value of particle density. Also, a velocity estimate of 10^7 cm/sec for a solar wind would require a downward revision of the particle density to some 10 cm^{-3} . Since the field for the furthestmost region was still decreasing, neither the interplanetary field nor its direction can be estimated. If the direction had been obtainable, a velocity might have been estimated utilizing Parker's model.

The value of the field rapidly drops between 7.8×10^4 and $9.39 \times 10^4 \text{ km}$, and at the outer

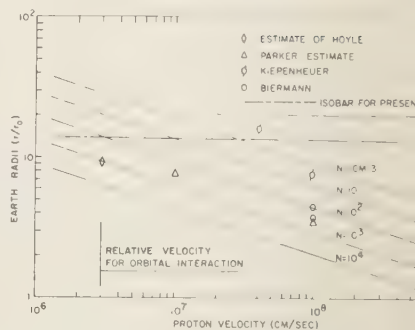


Fig. 13. The data of Pioneer I plotted to the expected geocentric cutoff in terms of relative plasma-earth velocity and plasma density, N .

end of the data run has dropped to $6 \times$ gauss. Current models of the transport of solar field outward [Gold, 1958; Parker, 1958] assure a radial field at many solar radii. It is to be taken as meaning a solar field transported outward primarily in the plane of the ecliptic. For relatively slow streaming velocity on the day of the flight, the field at the earth would have had to have a direction almost parallel to the orbit and thus be wound up at the sun. Since the value of field obtained is geometry-dependent, if the field streams radially outward from the sun and does not spiral, a small component has been measured. Reasonable estimates of corpuscular velocity, provided that a sufficiently dense gas existed, would require a field whose orientation was characterized by a large fraction of 90° in the vicinity of earth (i.e., directed about normally at distance). Then, almost the total field would be projected into the equatorial axis of the vehicle and measured by the magnetometer.

Small-scale disturbances. In addition to the phenomena reported above, there is a great variability in the data for region B having characteristic times of less than 1 second. Longer-term variations, shown in Figure 11, have not been spectrally analyzed, as the significant periods are of the order of data bursts. Therefore, these can only be determined by inspection of Figure 12. For the most part, two types of such disturbances have been observed, those that are clearly aperiodic, usually favoring temporary field decreases from ambient, and those having an almost per-

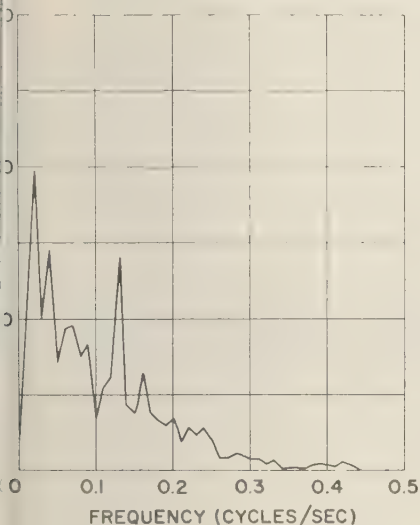


Fig. 14. The power spectral density of the small fluctuations of B_1 computed from the auto-correlation integral for the region of 8×10^4 km.

meter. This type of disturbance seems to be characteristic of all the data of region B, the mean power, however, being considerably higher at 93,800 km than over the prior part of the region. The variations are of large amplitude and are sufficiently complex so that it was not possible to calculate power spectral densities in order to determine the general character of these variations. A typical power spectral density is shown in Figure 14. It is seen that the spectrum consists of many discrete lines superimposed on a continuum varying approximately as ω^{-1} . Since the magnetometer was oriented so as to receive almost the total field, these fluctuations are primarily of a compressional type. They are perhaps 10 to 20 that additionally may have large changes in phase [Sonett and others, 1965] implying a rotation of the field in conjunction with, or separately from, amplitude changes.

It is assumed that these are traveling disturbances. Calculation of the Alfvén velocity $V = 10^2$ protons/cm³ yields a velocity $v \cong 10^2$ m/sec. For these velocities, the vehicle can be considered stationary. Then, utilizing the present cutoff of $f \sim 0.3$ cps noted in the power spectra for the region about 8×10^4 km, the upper bound on wavelength becomes $\lambda_c \sim 100$



Fig. 15. The radial component of vehicle velocity shown versus the geocentric distance for regions A and B, respectively.

km, and in the region of 9.3×10^4 km, $\lambda_c \sim 30$ km, since here $f_c \sim 1.0$ cps. If the plasma consists of electrons at 2 to 3 eV and protons of energy 5 keV (i.e., assuming an unmodified 10^3 km/sec solar stream), this tentative hydromagnetic activity has a minimum scale much less than the gyro radius for protons, this being $\rho_{H^+} \sim 10^3$ km, whereas for electrons, $\rho_{e^-} \sim 1$ km.

These disturbances do not seem to propagate into region A in appreciable amplitude. The marked dropoff in power density for the extreme radius is unexpected. Since the power spectrum extends to 1.0 cps for the farthest data, this would require a decrease of particle density of about a factor of 10 to account for the wavelength, assuming Alfvén propagation and a nonchanging field. The rapid change in power density might be attributed to the production of longitudinal oscillations for which the data reduction has not been completed. If such a process does take place, hydromagnetic energy will be propagated in the polar directions. Thus, a means would be available for a decrease in the power density as a function of distance.

It is entirely possible that these fluctuations represent quasi-stationary magnetic discontinuities except for those that are clearly oscillatory upon visual inspection of the record. This would obviate the difficulties inherent in assuming traveling disturbances. Assuming this latter model, the power spectra that have been computed form a measure of the scale. However, motion of the plasma will modify the values.

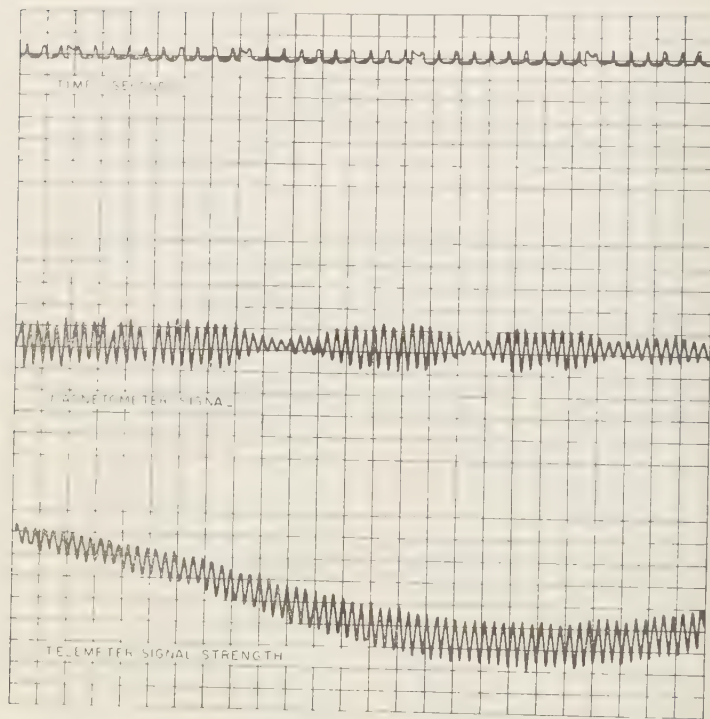


Fig. 16. Characteristic quasiperiodic oscillations of B_L occurring in region B. See Figure 17 for amplitude scale.

For the data shown in Figure 15, since the vehicle velocity is ~ 1 km/sec (see Fig. 15), 1.0 cps corresponds to a distance of 2.2 km. The cutoff frequency is about 0.3 cps. Thus the smallest scale would be about 7.5 km. The apparent scale would increase with any radially inward motion of the gas as the velocity relative to the vehicle would increase. Assuming that the structure noted represents the pile-up of hydro-magnetic material, it would be easier to account for the phenomena on the basis of a solar wind rather than on the basis of an interaction of the earth with a stationary plasma, for reasons previously given concerning shear effects. If, in fact, this model is correct, it might imply a sweeping-out of gas by the earth, though it is difficult to see how orbital motion could cause this as the data were taken at approximately local noon, some 90° away from the stagnation point (see Fig. 11). The power spectra allow the magnitude of the variations to be determined. For example, the mean value of $(\Delta B)^2/\Delta\omega$ is

5×10^{-7} gauss²/cps over a band of about 0.1 cps. Thus, $\Delta B \sim 3 \times 10^{-4}$ gauss. This represents a $\Delta B/B$ of 1.0 to 1.5.

Assuming that dissipative effects are absent the propagation of these disturbances in region A would require that they be reduced in magnitude to values comparable with the 10 to 15 per cent variability noted in the data in region A. This type of argument follows from consideration of the ratio of magnetic-disturbance energy density and ambient-field energy density, from which it would be expected that inward-traveling disturbances would show a magnitude of decrease, $\Delta B/B$, varying as the third power of the radius.

In attempting to test for propagation earthward of these disturbances, the time-annotated record has been compared with surface observations on the day of the flight, these having been taken at Borrego, California, by W. Campbell of the UCLA Institute of Geophysics. Owing to the extremely low magnetic activity

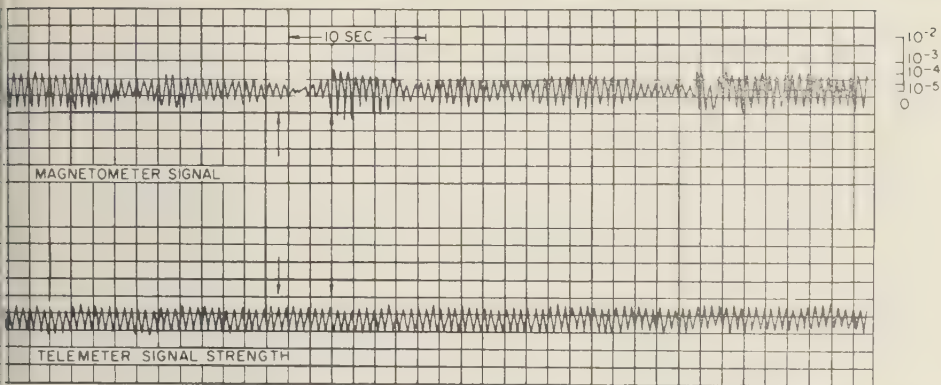


Fig. 17. A typical section of magnetometer record for region B. These data contain phase changes and abrupt changes in amplitude. The amplitude fluctuations are considerably accentuated in this figure because of AGC effects in the magnetometer amplifier.

outstanding events could be noted, so that correlation was not possible. Of course, reflection in the ionosphere [Dessler, 1958] would possibly distort the wave form and introduce a serious dispersion problem complicating any comparison.

Two samples of the data from region B have been included in Figures 16 and 17, respectively, to demonstrate two examples of the structure that appears to be present. Figure 16 shows what appears to be a hydromagnetic disturbance of amplitude $\Delta B/B \sim 10$ and with varying period. Figure 17, on the other hand, displays the more characteristic structure in this region associated with some phase shifting.

Conclusions. This experiment has provided the first direct data concerning the physical state of the distant geomagnetic field. Since these data were taken on a magnetically quiet day, they may provide a lower bound for measures of transient activity. No large-scale ring-type convection was found in the region of some 10,000 to 44,000 km, though there is some evidence for small-scale convection processes. The vestigial field (78,000 to 93,000 km) displayed a cutoff at 13.6 radii. Assuming interaction of the earth with a stationary plasma, the particle density, if protons, was approximately 100 cm^{-3} . The lowest value of field measured centered about 5×10^{-6} gauss with indications that the field was continuing to drop. Thus, an upper bound has been established for the interplanetary field, provided that the field was extended

radially outward from the sun and strong spiraling was present.

Also, either strong hydromagnetic traveling disturbances or a considerable fine structure in the vestigial field has been found.

Note added in proof: Subsequent analysis has tended to indicate that many of the disturbances discussed above are sharp crested waves with rise time of, at most, a few seconds, followed by much slower decay implying earthward progressing nonlinear or shock type disturbances.

Acknowledgments. This experiment represented the combined and sustained efforts of perhaps thousands of individuals in order to meet a time scale of only a few months from inception of the program to the flight. It is impossible to acknowledge directly the considerable help and efforts of so many people.

The Program Director was G. E. Mueller, Associate Director of the Research and Development Division of Space Technology Laboratories; the Project Engineer for the fourth stage (payload) was A. W. Newberry, under whose direct supervision the payloads were constructed. The amplifier was designed by M. G. McLeod.

We wish to acknowledge the considerable help of I. Abrams in processing the data, of Professor A. C. B. Lovell in tracking the Pioneer I vehicle, and of W. Campbell of the UCLA Institute of Geophysics in providing data on geomagnetic disturbances. We wish to express our gratitude to these individuals for their support in carrying out this experiment.

REFERENCES

- Alfvén, H. O. G., Interplanetary magnetic field, in *Electromagnetic Phenomena in Cosmical Physics*, edited by B. Lehnert, Cambridge University

- Press, pp. 284-290, 1958 (Symposium 6 of the IAU).
- Behr, A., and H. Siedentopf, *Z. Astrophys.*, **32**, 19, 1953.
- Biermann, L., *Observatory*, **77** (898), 109-110, 1957.
- Dessler, A. J., *J. Geophys. Research*, **63**, 507-511, 1958.
- Dungey, J. W., *Cosmic Electrodynamics*, Cambridge University Press, 1958.
- Gold, T., The magnetic field in the corona, in *Electromagnetic Phenomena in Cosmical Physics*, edited by B. Lehnert, Cambridge University Press, pp. 275-280, 1958 (Symposium 6 of the IAU).
- Goldstein, H., *Classical Mechanics*, Addison-Wesley Press, Cambridge, Mass., 1951.
- Hoyle, F., *Phys. Rev.*, **104**, 269, 1956.
- Judge, D. L., M. G. McLeod, and A. R. Sims, A high-sensitivity transistorized search-coil magnetometer, submitted to *Rev. Sci. Instr.*, 1959.
- Kiepenheuer, K. O., *Z. Naturforsch.*, **6a**, 627, 1951.
- Kiepenheuer, K. O., Solar activity, in *The Sun*, edited by G. P. Kuiper, University of Chicago Press, pp. 322-465, 1953.
- Margolis, J. S., Calibration of a magnetometer coil by the method of spinning in a uniform magnetic field, *GM 447-177*, June 19, 1958, Space Technology Laboratories, unpublished.
- Morrison, P., *Phys. Rev.*, **101**, 1397-1404, 1956.
- Parker, E. N., *Lockheed Symposium on Magnetohydrodynamics*, Stanford University Press, 1958a.
- Parker, E. N., *Physics of Fluids*, **1**, 171-187, 1958b.
- Rosen, A., C. P. Sonett, and P. J. Coleman, Ionizing radiation at altitudes of 3500 to 36,000 kilometers: Pioneer I, *J. Geophys. Research*, **709-712**, 1959.
- Sims, A. R., Effect of free precession and varying magnetic field upon the magnetometer signal and the aspect indicator, *TM-59-351*, April 1959, Space Technology Laboratories, unpublished.
- Sonett, C. P., S. C. Baker, P. J. Coleman, D. L. Judge, J. M. Kelso, M. G. McLeod, and A. Rosen, Scientific instrumentation of Pioneer I and II space probes, submitted to *Rev. Instr.*, 1959a.
- Sonett, C. P., D. L. Judge, and J. M. Kelso, Evidence concerning instabilities of the distant geomagnetic field: Pioneer I, *J. Geophys. Research*, **64**, 941-943, 1959b.
- Stratton, J. A., *Electromagnetic Theory*, McGraw-Hill Book Co., New York, 1941.
- van de Hulst, H. C., *B.A.N.*, **10**, 325, 1950.
- Vestine, E. H., *J. Geophys. Research*, **58**, 127-133, 1953.

(Manuscript received August 22, 1959; revised November 2, 1959.)

Some Characteristics of the Upper-Air Magnetic Field and Ionospheric Currents¹

ALFRED J. ZMUDA

*Applied Physics Laboratory
The Johns Hopkins University
Silver Spring, Maryland*

Abstract. Characteristics of the upper-air magnetic field and ionospheric currents are determined through an analysis of published rocket data on the magnetic scalar intensity. For the region between the earth's surface and the *E* layer of the ionosphere, the observed values are compared with values obtained by extrapolating the surface vector field. The agreement between the two sets of values is very good for equatorial flights but only fair for a flight at White Sands, New Mexico. The equatorial, ionospheric current density, which has a maximum of about 21 amp/km², varies considerably with time and location of the rocket flight. The equatorial electrojet has a current intensity of about 130 amp/km and flows practically along the magnetic equator. Some of the computed currents associated with the normal magnetic daily variation are at variance with those expected from considerations, such as harmonic analysis, of the surface transient field and sheet-current approximations. In the area around White Sands (geomagnetic latitude 41°N) a large negative magnetic anomaly exists that may contribute to the formation of the region of low intensity of radiation that lies between the two Van Allen radiation belts.

Introduction. Upper-air magnetic data supplement surface information in analyses treating fundamental geomagnetic problems as the description of the field for regions above the earth's surface and the characteristics of the external currents that produce the transient magnetic variations.

The most general description of the geomagnetic field is that determined through the application of spherical harmonics to the vector magnetic intensity. Beginning with Gauss, a number of authors performed harmonic analyses of the surface data which in general, fit the observed surface field to within a few per cent and indicate that the practically time-independent part of the field stems from sources inside the earth; see for example *Chapman and Bartels* [1940] and *Vestine* [1947]. Whether there is a superposed source of external origin is a question currently being discussed in relation to the geographic distribution of cosmic-ray intensities and cut-off momenta [*Simpson and others*, 1956; *Katz and others*, 1958; *Rothwell*, 1958; *Queenby and Webber*, 1959].

Superposed on the steady part of the geomagnetic field is the transient field whose principal variations are classed as *S* (solar), *L* (lunar), and *D* (disturbance). The solar and lunar variations, exhibiting a moderately regular daily variation, relate to ionospheric currents in large measure set up by dynamo action. [See the discussion of the Stewart-Schuster theory by *Chapman and Bartels*, 1940.] In the equatorial region the normal *S* and *L* currents are augmented by an electrojet current which is confined to a narrow band [*Chapman*, 1951; *McNish*, 1937; *Baker and Martyn*, 1952; *Alexander and Onwumechilli*, 1957; *Egedal*, 1947]. The *D* field stems from currents composed of charged particles emitted from the sun. The location of these currents is at present unknown, but some of their characteristics have been mapped by *Vestine* [1947].

Although many properties of the main and transient geomagnetic field are determinable through analysis of surface values, the need for upper-air magnetic investigations has long been recognized [*Chapman*, 1954; *Vestine*, 1956]; and rocket experiments were performed by *Maple and others* [1950]; *Singer and others* [1951]; *Heppner and others* [1958]; and *Cahill* [1959a, b].

It is the purpose here to connect, whenever

¹This work was supported by Air Force Cambridge Research Center, Geophysics Research Directorate, Bedford, Massachusetts.

possible, the upper-air values of the scalar magnetic intensity, the only magnetic element measured in the rocket flights, with surface data on the vector magnetic intensity and with the vector current density producing the daily magnetic variations. In the subsequent calculations and the interpretation of the results, a number of simplifications are possible. To present these approximations in their proper perspective, an attempt is made to discuss in sufficient detail the mathematical background and the steps in the computations. When currents are discussed, a volume distribution is treated, so that the pertinent material may add, at least a little, to existing sheet-current treatments.

Some aspects of the physical and mathematical background. The geomagnetic field may be considered here, where secular changes are neglected, as consisting of two general parts: a time-independent and a transient field. For regions at or near the earth's surface and for regions outside the sources of magnetism, both types of field are derivable from a potential V that satisfies Laplace's equation $\nabla^2 V = 0$ (where ∇^2 is the well known Laplacian). The most recent, for 1955.0, harmonic analysis of the time-independent surface field was performed by *Finch and Leaton* [1957], who computed 48 coefficients of the part of the harmonic series that relates to the field of internal origin. The steady field described by the finite harmonic representation, hereafter called the main field, generally represents the total observed surface field to within a few per cent. The differences between observed and computed surface values stem either from anomalous fields or from small errors in the harmonic coefficients deriving from a lack of data for many sections of the earth's surface. Anomalous fields are used here in the sense that they are time-independent fields not represented by the terms contained in the finite expansion of harmonics with the coefficients computed by *Finch and Leaton* [1957].

Spherical harmonic analysis shows that the surface transient field consists of two parts: a primary field of origin external to the earth, and a secondary field due to currents inside the earth but induced by the primary field. At the surface the secondary field represents about a third to a fourth of the total transient field. Given an independent determination of the location of the external currents, one can com-

pute through spherical harmonics of surface data the intensity of the so-called toroidal current which may be only part of the total ionospheric currents, as will be later discussed.

Although harmonic analysis is the most general approach in geomagnetic interpretation it is sometimes convenient (for example, in the treatment of localized effects such as those due to the equatorial electrojet) to use, in place of spherical harmonics, a Taylor series [*Zmuda*, 1956] in which the derivatives in the vertical direction are obtained from surface values of the components; the Taylor series is used here to extrapolate the components of the transient field from the earth's surface to the base of the upper-air current system.

For points inside the currents the mathematical considerations differ in many respects from those governing the behavior of the field in regions below the currents. The magnetic intensity connected to the currents, say \mathbf{F}_c , satisfies the relationships, in electromagnetic units,

$$\text{and} \quad \left. \begin{aligned} \nabla \times \mathbf{F}_c &= 4\pi \mathbf{i} \\ \nabla \cdot \mathbf{F}_c &= 0 \end{aligned} \right\} \quad (1)$$

where \mathbf{i} is the current density and where the displacement current is neglected. To the solutions of (1) are added those derived from the solution of Laplace's equation, in order to describe the geomagnetic field of origin internal to the earth.

The general solutions of (1) are of two types [see, for example, *Bullard and Gellman*, 1954, or *Elsasser*, 1947]: toroidal and poloidal magnetic fields connecting respectively to poloidal and toroidal currents. The toroidal magnetic field is confined entirely within the current system and is zero at the boundaries; the characteristic of this field, and of the current producing it cannot be determined from observations taken outside the region containing the currents. The poloidal magnetic field connects to the field existing outside the current region through the application of the boundary condition that all components of magnetic intensity are continuous on passage across the boundaries. In general, for points inside the currents, the radial variations of all magnetic elements differ from those occurring outside the currents, so that the boundaries of the currents are marked by points where a change in radial dependence is found.

in place of the general series solution of (1) use here, for the transient field, the differential form of $\nabla \times \mathbf{F} = 4\pi\mathbf{i}$, which is in component form

$$\left. \begin{aligned} i_\theta &= \frac{\partial Y}{\partial r} + \frac{Y}{r} + \frac{1}{r \sin \theta} \frac{\partial Z}{\partial \lambda} \\ i_\lambda &= -\frac{\partial X}{\partial r} - \frac{X}{r} + \frac{1}{r} \frac{\partial Z}{\partial \theta} \\ i_r &= -\frac{1}{r \sin \theta} \left[\frac{\partial}{\partial \theta} (Y \sin \theta) + \frac{\partial X}{\partial \lambda} \right] \end{aligned} \right\} \quad (2)$$

where r , θ , and λ are the spherical polar coordinates representing respectively the distance from the earth's center, the colatitude, and east longitude; X , Y , and Z represent respectively the northern, eastern, and downward components of the magnetic intensity. The components of the current refer to the components of the total current, toroidal and poloidal. The vertical component can be neglected here, where points far the boundaries of the current shell are to be treated, on two related counts. First, a boundary condition is that i_r vanishes at the boundary; second, any vertical current initially flowing inside the shell ultimately leads to a polarization of the shell [Chapman and Bartels, 1940] and hence to a curtailment of the vertical current flow.

Although the components of the vector magnetic intensity, as well as the magnitude of the scalar intensity, are continuous functions in all regions above the earth's surface, it is sometimes convenient to interpret the radial changes around in passing through current systems as discontinuities encountered in crossing a thin, uniform current sheet of infinite horizontal extent but of zero thickness. In this case [see for example Chapman, 1954], the current intensity (sometimes called also the integrated current density or current per unit, horizontal line width) flowing in the current sheet is given by

$$2\pi j = H_s \quad (3)$$

where H_s is the surface value of the horizontal, and sole, component of the intensity of the transient field. In this approximation, the transient intensity is oppositely directed on the two sides of the unbounded plane and has a magnitude that is invariant with altitude. The current intensity j is related to the current

density i through

$$j = \int_0^h i \, dh \quad (4)$$

where h is the radial thickness of the current system.

If H_s in (3) is replaced by one-half the difference in the values of H found at points immediately below and above the current system, the resultant j is very nearly equal to that which would flow in a spherical sheet of, mathematically speaking, zero thickness. In both sheet approximations, laterally unbounded and spherical, the direction of current flow is normal to the direction of the horizontal component of the transient field and is, according to the well known, right-hand rule, eastward when \mathbf{H} points northward in regions below the current system and southward in regions above the current system; when these directions of \mathbf{H} are interchanged, the current flow is from east to west.

Values of current intensity j have been computed by McNish [1937] and by Bartels using coefficients determined in harmonic analysis of the surface data on the daily magnetic variation. Values obtained in this manner refer only to the toroidal ionospheric currents, which, as previously noted, produce the related poloidal magnetic field. The poloidal currents cannot be determined through analysis, harmonic or otherwise, of surface data alone, since, as also was previously noted, these currents produce a field that is confined entirely within the region containing the currents and, hence, a field that is not observable, for example, at the earth's surface.

Analysis and interpretation of rocket data. The analytic method followed here and the interpretation of the results will now be discussed in detail for the data of rocket flight SUI-86 and in general for the other rocket flights treated. The rocket SUI-86 was fired by Cahill [1959a, b] on October 18, 1957, in the equatorial region, near a set of magnetic observatories maintained by the Scripps Institution of Oceanography. The flight data on the magnetic scalar intensity, shown in part in Figure 1, have the following characteristics. At altitudes below 96 km, the radial variation in intensity has a practically constant slope, following closely an inverse-cube law and indicating the predominance of the dipole part of the main field. At about 96 km,

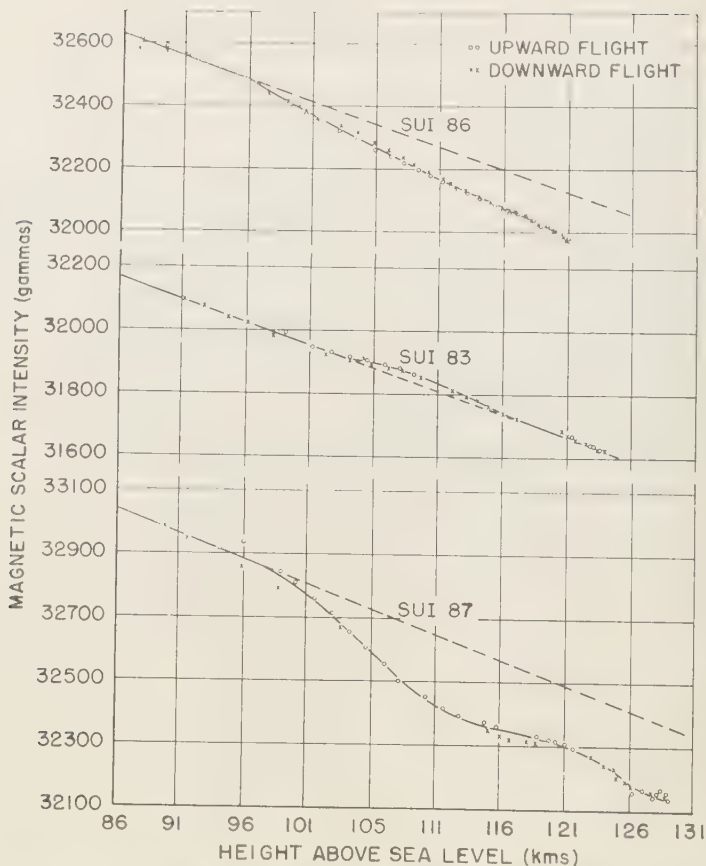


Fig. 1. Variation with altitude of scalar magnetic intensity, for some rockoon flights in the equatorial regions. (Data points are reproductions of those published by Cahill.)

the abrupt change in the slope of the intensity curve marks the point of entry into the current system. The rocket probably did not pass completely through the current system, since, at the peak altitude of 122 km, the slope is not nearly equal to that occurring for points below the currents. The measured set of scalar intensities contains contributions from a number of magnetic sources, with fields classifiable, in general, as steady and time-dependent fields of origin internal as well as external to the earth. In the separation of the rocket scalar values into their respective parts, surface vector data are a considerable aid.

The curves for the daily magnetic variation at the Scripps observatories are shown in Figure

2; the diurnal variation at the launching spot, at $0^{\circ}02'N$ and $160^{\circ}52'W$, is here assumed to be that which occurred at the Jarvis Observatory located about 60 miles to the southeast. At Jarvis, the abnormally large range of the daily variation in the horizontal component H characterizes the changes in H found at observatories near the magnetic equator, where $I = 0$. Chapman [1951] attributes the abnormality, which has a maximum effect at about 1100 local time, to the field of a concentrated current (which he named the equatorial electrojet) that is superposed on the field of the currents producing the normal daily variation. For the time of the rocket firing, 1356, 165 MT, the H value at Jarvis, as well as the values of the other

TABLE 1. Location and Magnetic Properties of Scripps Observatories*

Station	Latitude	West Longitude	Horizontal Force, H, γ	Declination, D , East	Dip, I , North
Jarvis	0°23'S	160°03'	34, 507	9°16'	2°12'
Fanning	3°54'N	159°24'	32, 978	9°28'	10°34'
Palmyra	5°53'N	162°05'	32, 424	9°47'	13°05'

* With the exception of the I values for Fanning and Palmyra, the listed magnetic data are the mean values for October 18, 1957.

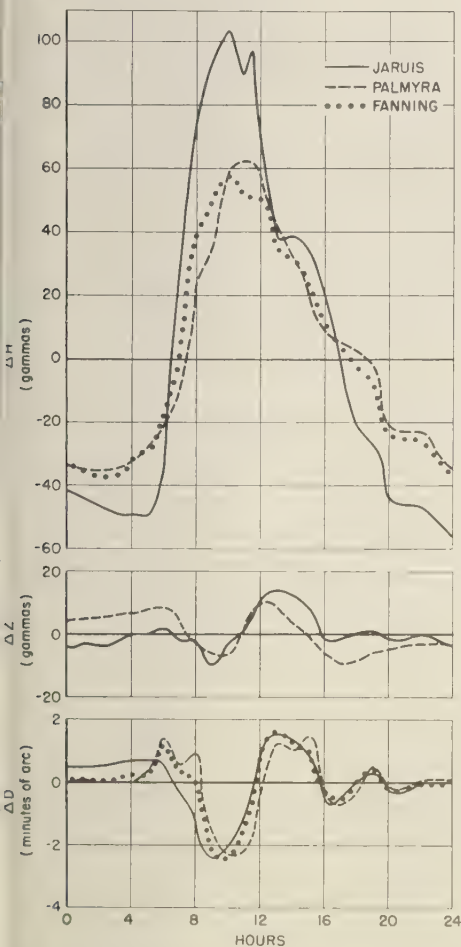


Fig. 2. Daily magnetic variation at the Scripps observatories, for 165 MT, October 18, 1957. The ordinates are the departures of the elements from their daily mean value. The rocket SUI-86 was launched at 1356, 165 MT.

elements, differs only slightly from the corresponding value at the other observatories, indicating that during the flight of this rocket the effects of the equatorial electrojet are probably small, if not negligible.

In the interpretation of the observatory magnetograms, a question arises about what part of the total surface field relates to the field, primary and induced, of the ionospheric current system. If the deviations from the daily mean value are used, it is implied that the currents producing the field of the daily variation reverse direction during the length of day, since the currents producing positive deviations from the mean will be opposite in direction to those producing negative deviations. An alternative interpretation is that the temporal variations are always positive departures taken with respect to the daily minimum value. In this case there will be no diurnal reversal of current flow, but the currents will vanish for some night-time period. (This latter viewpoint is in line with, but may go a little farther than, the suggestion of Chapman [1951] that the noon tropical current may correspond to the daytime excess of H over the night-time value rather than to the departure from the daily mean.) Both interpretations will be considered here. For the time of the rocket flight, the Jarvis values of the transient components X , Y , and Z are respectively 36, 19, and 12 γ , when deviations from the mean are used; and 87, 32, and 22 γ , with deviations from the minimum value ($1 \gamma = 10^{-6}$ gauss).

For points either below or at the lower boundary of the overhead current system, the transient vector field may be computed through the application of the Taylor series to the surface values at the three observatories. These calculations showed that the flight-time transient field in the Jarvis area is practically invariant with

TABLE 2. Scalar Intensities for SUI-86 for Points below the Ionospheric Currents, gammas

Height above Sea Level, km	Measured Values	Values Computed through Vector Addition of Main and Transient Field			
		Transient Field Used as Devia- tion from Daily Minimum Value	Difference be- tween Observed and Computed Values	Transient Field Used as Devia- tion from Daily Mean Value	Difference be- tween Observed and Computed Values
40	33,367	33,300	67	33,248	119
60	33,036	32,983	53	32,930	106
80	32,717	32,670	47	32,618	99
96	32,486	32,423	63	32,371	115

altitude: a maximum difference of about 5 γ existed between the surface value and that for the altitude of 96 km. This invariance is qualitatively suggested by the closeness of the time-of-flight values of the transient elements at the observatories (see Fig. 2); the similarity indicates practically zero values of the surface derivatives used to effect the upward continuation of the transient field. As previously noted, the main field is extrapolated by means of the harmonic coefficients determined by *Finch and Leaton* [1957]. For points below the currents, the scalar intensity obtained through the vector addition of the transient and main field is shown in Table 2, along with some observed values. The observed data are average values interpolated from the upward and downward flight measurements kindly sent, along with data for the other SUI flights here discussed, to the author by Cahill, who estimates that $\pm 5 \gamma$ is the maximum relative error between the measured values. (The absolute error is difficult to estimate in the existing rocket data.)

Good agreement exists between the measured values and either of the two sets of computed values, although the agreement is better for the set in which the surface transient field is treated as a deviation from a minimum value. Considering the possible errors of measurement, the difference between the observed and computed values is practically constant with altitude. Now, since the main field of internal origin and the total transient field is accounted for, a constant difference can be attributed to a steady field of external origin, such as a ring current, whose field values near the earth's surface are principally represented by harmonics of the first degree. This possibility is further considered in relation to the results of other rocket flights. In general,

in view of the form of the harmonic solutions the variation of the difference with altitude permits at least a qualitative description of some properties of the field responsible for the difference. If the difference field decreases with increasing elevation, it will probably be associated with sources inside the earth, since the contribution of the source will decrease with increasing distance. On the other hand, a constant difference or a difference increasing with increasing altitude will probably be associated with a field of an external source.

Before considering the field for points inside the current system, it is desirable to associate the scalar differences found in Table 2 with the vector components of an anomalous time-independent field. In the equatorial regions where $\mathbf{X} \simeq \mathbf{F}$ and where doubling the known Y and Z values would increase the scalar intensity by only a few gammas, no serious error is introduced by considering the difference values as values of the X component. (This procedure is not valid for the nonequatorial region; a useful approach for it is treated later.) From the variation with altitude of the anomalous field for points below the current system, the anomaly values at points inside the current system are determined.

For the current region there are available measured values of the total scalar intensity and, based on previously treated considerations, computed values of the time-independent vector field. With this information, the characteristics of the transient field and its associated vector current density can be determined for points in the current region. The application of the following characteristics of the current and the transient field assists in this part of the analysis (1) The vertical component of the current

vanishes at the boundaries of the current system and, hence, may be neglected in the regions near the bounding surfaces. (2) For points on the lower boundary, the components may be computed from the surface values. (3) With reference to equation 2, the vertical component of the field relates to the current through its derivatives in the horizontal directions. For the time of the night, these derivatives are practically zero at the earth's surface, at the lower boundary of the currents, and, since the components are continuous in crossing the boundary, also at points located inside the current system but not far removed from the boundary. These horizontal derivatives can be neglected, as can the terms Y/r and X/r ; the current density associated with all these factors is, at most, of the order of 10^{-13} emu/cm² (or 10^{-2} amp/km²). Under these conditions, equation 2 becomes

$$i_\theta = \frac{1}{4\pi} \frac{\partial Y_t}{\partial r} \quad \text{and} \quad i_\lambda = -\frac{1}{4\pi} \frac{\partial X_t}{\partial r} \quad (5)$$

r , in terms of the horizontal component of the transient field, after a proper rotation of the coordinate system,

$$i = -\frac{1}{4\pi} \frac{\partial H_t}{\partial r} \quad (6)$$

where i is the current density flowing in the direction of the normal with respect to H_t and the subscript t identifies the transient field.

The direction H_t often differs considerably from the direction of the total horizontal field, which in addition to the transient field contains as the predominantly major portion) a part due to the time-independent field. Thus, while the D variogram (see Fig. 2) shows that the directional changes of total H are small, it should be noted that these small changes are due to large changes in the direction and magnitude of the transient field. If the direction of current flow is taken as normal to the direction of the total horizontal field, which scarcely changes with time, large errors result in the computed direction of flow.

With reference to (5), thus to determine the current density, the derivatives in the vertical direction of X_t and Y_t must be extracted from the available data. A useful characteristic applicable at this stage is that the functional forms of the r dependence for X_t and Y_t will be practically identical, since, among other

considerations, the radial variations for these elements are identical in the general term of the solutions of (1). Consequently, to first order, at least,

$$\frac{X_{th}}{X_{t_0}} = \frac{Y_{th}}{Y_{t_0}} \quad (7)$$

where h is the height above the lower base of the current system. The values of the transient components at the lower boundary, X_{t_0} and Y_{t_0} , respectively, are computable from the data obtained at the earth's surface.

Using equation 7 and the expression

$$F = [X_t^2 + Y_t^2 + Z_t^2]^{1/2}$$

and neglecting Z_t which is the smallest component and which (as previously noted) relates to negligible currents, one obtains for X_t at each altitude, for points inside the current system,

$$X_t = (-b + \sqrt{b^2 - 4ac})/2a \quad (8)$$

where

$$a = 1 + \alpha^2,$$

$$\alpha = Y_{t_0}/X_{t_0},$$

$$b = 2X_{mf} + 2X_A + 2\alpha Y_{mf},$$

$$c = Z_{mf}^2 + (X_{mf} + X_A)^2 + Y_{mf}^2 - F^2,$$

and where the subscript mf refers to the main field and A to the anomaly field. For each value of X_t a value of Y_t is computed from (7).

Equation 8 represents one root of a quadratic expression for X_t ; the second root is discarded because its value exceeds the value of the total scalar intensity. From (7) and (8) the values of the transient components for points inside the current system are obtained; with these values and (5), the components of the current density are determined.

The variation with altitude of the amplitude of the current density (see Fig. 3) shows the existence of two distinct shells of current. In the lower layer, extending approximately from 97 to 110 km, the current decreases approximately linearly from the maximum, which occurs at the lower boundary of the shell. In the upper layer, extending approximately from 110 to 121 km, the current density is practically constant. The direction of current flow is clockwise normal to the direction of the surface transient field; this direction makes a clockwise angle of

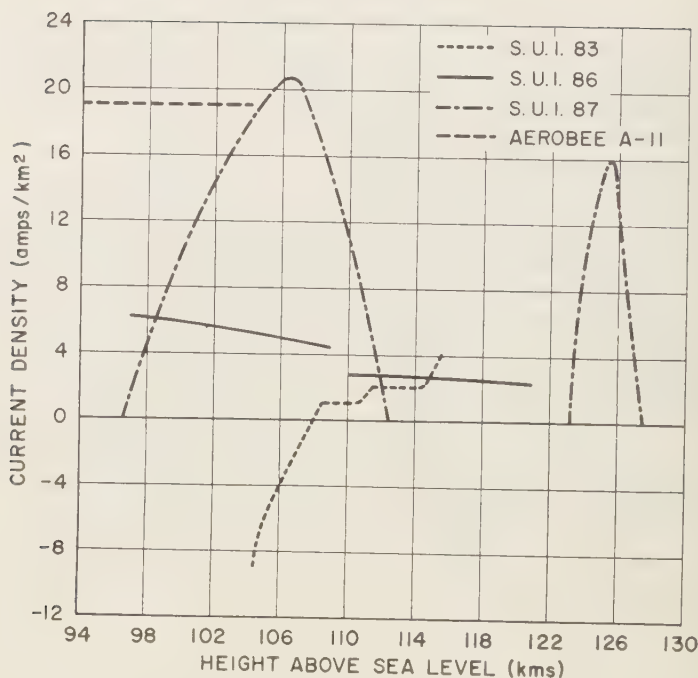


Fig. 3. Magnitude of current density for equatorial flights.

about 20° with the geographic west-to-east direction, or about 101° with the direction of the total (time-independent and transient) horizontal force, whose declination is 9°E . The current intensity, or integrated current density, is 96 amp/km. All these current characteristics are equally applicable in the two different interpretations of the transient field: where the surface transient field is considered as a deviation (1) from a minimum value and (2) from a mean value. The reason for the ambiguity lies in the fact that the current density depends primarily not on the absolute values of the transient field, which were different in the two cases, but on the vertical derivatives of this field, which were practically identical. (The ambiguity exists also in the other rocket flights discussed here.) The resolution of the dilemma may result from the analysis of night flights: for the night-time, the ionospheric currents vanish in one interpretation (using deviations from a minimum value) and reverse direction in the other.

Inside the current system, H_i decreases with increasing altitude, a characteristic also evident

in Figure 1, where, after the rocket enters the current system, the total intensity decreases more rapidly, owing to the decrease in H_i . For points located above the current system, H_i also decreases, in theory, with increasing elevation, and it is necessary to distinguish the decrease from that connected with the current through (6), and in the process mark the upper boundary of the current system.

For points above the current system, the harmonic description of the transient field contains the radial factors $(b/r)^{n+2}$, where b is the radial distance from the upper boundary of the current system to the center of the earth and n is the degree of the spherical harmonic. The degree n , as well as the order m , of the harmonic will be the same as in the description of the transient field at the earth's surface. From harmonic analysis of the surface S field, n has maximum values of about 8; for a harmonic description of the field of the electrojet, harmonics of higher degree are probably necessary. It is unlikely, however, that the required harmonics are of sufficiently high degree to cause

an appreciable reduction in H_t over the small radial distances treated here. For example, suppose that $H_t = 100 \gamma$ at the upper boundary of a current system located about 100 km above the earth's surface (or about 6500 km from the earth's center). If $n = 60$, the value of H_t at a point 10 km above the current system is 90γ ; the reduction in absolute value is small. Thus, for practical purposes, a radial constancy of H_t (or equivalently in X_t and Y_t) marks the region near but above the current system, a result coinciding with that deducible from (6), where the current vanishes when $\partial H_t / \partial r = 0$. By this criterion currents still exist at elevations around 122 km, the maximum elevation achieved in this flight.) This constant value of H_t above the current system differs from that below the currents. Consequently, the slopes of the total intensity curve differ slightly in the two regions outside the current system. The type of difference occurring depends on the value of H_t at the two boundaries and the manner in which H_t varies inside the current system.

In this, or any, equatorial flight where both the transient and total (main, anomalous, and transient) fields are principally horizontal, though not necessarily in the same direction, the current density may sometimes be estimated from

$$i = -(1/4\pi)(\partial \Delta F / \partial r) \quad (9)$$

where ΔF , a value of scalar intensity assumed to be that of the transient field, is the difference in scalar intensity between the observed value and the value obtained from the dashed part of the curve in Figure 1. This dashed curve is the continuation, without a change in slope, of the intensity values from the region below the currents to the region inside the currents. For SUI-86, the amplitude of the current ob-

tained with (9) is practically identical with that shown in Figure 3, which is obtained using the previously discussed vector and scalar relationships. The direction of current flow, which is now, continuing the approximation of (9), along the clockwise normal to the direction of the total intensity, is, however, in error, making a clockwise angle of 9° with the geographic eastern direction, whereas before this angle was computed to be 20° . For other equatorial flights, it may be that (9) will yield erroneous results also in the amplitude of the current; this error depends in general on the relative direction of the transient and total horizontal intensity, a relative direction that varies during the course of the day. In nonequatorial regions equation 9 is definitely not applicable because the transient and total field are no longer in the same general direction; and one must use, for example, the other approach followed here, which in addition to yielding the characteristics of the current gives information on the vector geomagnetic field.

In addition to rocket flight SUI-86, the following were analyzed: SUI-83 and 87, Aerobee A-11, and Aerobee NRL-33. The additional Iowa flights were analyzed in practically the same manner as SUI-86; with the Aerobee flights, because of the varying amounts of available information, some of the specific characteristics in the analysis differed from those previously discussed.

The main field for each flight except Aerobee A-11 was computed by means of the harmonic coefficients published by *Finch and Leaton* [1957], for the rocket launch point and for points spaced along the entire trajectory. For the Iowa flights, the transient field values were taken to be those occurring at the nearest observatory at the time of the rocket flight. For

TABLE 3. Some Characteristics of the Rocket Flights

Flight Designation	Geographic Coordinates		Time of Firing, MT	Date of Firing	Maximum Altitude, km
	Latitude	West Longitude			
SUI-83	$3^\circ 23'N$	$158^\circ 41'$	1359 165	10/17/57	127
SUI-86	$0^\circ 02'N$	$160^\circ 52'$	1356 165	10/18/57	122
SUI-87	$2^\circ 05'S$	$161^\circ 05'$	0907 165	10/19/57	129
Aerobee A-11	$11^\circ 06'S$	$88^\circ 27'$	1120 90	3/22/51	105
Aerobee NRL-33	$32^\circ 24'N$	$106^\circ 23'$	0052 105	7/5/56	162

SUI-S3, the observatory was Fanning, located about 60 miles northwest of the launch point; the transient field values are $X_t = 72 \gamma$ and $Y_t = 29 \gamma$ (the Z_t values were not available) when deviations from the minimum value are used, the only case reported for this flight as well as for SUI-87. The rocket SUI-87 was launched at a point about 140 miles southwest of Jarvis, where the time-of-flight transient values were $X_t = 220 \gamma$, $Y_t = 14 \gamma$, and $Z_t = 2 \gamma$. The data obtained in the pioneer Aerobee experiments of *Singer and others* [1951, 1952] were measured with respect to an undetermined surface value; therefore, it is not possible to compare these values with any computed field. For the flight A-11, however, which penetrated the equatorial ionospheric currents, some pertinent characteristics of the currents may be estimated in the following manner.

For the purpose at hand, it is informative to recall the following flight characteristics of A-11, in addition to those listed in Table 3. The launching took place at a point about 3° south of the dip equator; the time of firing closely corresponded to the time of peak overhead current flow. A decrease in scalar intensity of 1300γ was measured as the rocket moved from sea level to a height of 93 km, the altitude at which the currents were encountered; inside the current system, the decrease was $400 \pm 50 \gamma$ for the altitude range 93 to 105 km. The magnetic data indicate that the rocket had not passed completely through the current system and that an anomalous field may exist in the region below about 50 km.

The author estimates that, for points inside the currents, a total decrease in intensity of about 280γ relates to the field of the currents. In obtaining this value, it is assumed (1) that the horizontal force equals the total force and (2) that for the time-independent part of the field the radial variation of the intensity for points inside the currents equals the average radial variation, about $10 \gamma/\text{km}$, of the total intensity for points below the currents but between 50 and 93 km. With the decrease of 280γ over an altitude of 12 km, the average current density becomes, using (9), around $19 \text{ amp}/\text{km}^2$, and the current intensity $228 \text{ amp}/\text{km}$. This current density is considerably larger than values computed for the Jarvis area, although

the current intensity is not out of accord. The author's estimate of the current is considerably less than that of *Singer and others* [1952]; the respective values are in the ratio of 280 to 400. These authors incorrectly associated the entire intensity change of 400γ with the field of the currents.

An additional set of data has been published by *Heppner and others* [1958] for a night-time flight at White Sands, New Mexico. No currents were detected on this flight, which reached a maximum altitude of 162 km; the observed values are here compared with those computed for the main field. No near-by observatory data were available; but if the data from the Tucson observatory were used, the constant value of 10γ would be added to the computed main field values. The Tucson records indicate that the day of this flight was magnetically quiet.

Discussion of results. For points below the current system there is very good agreement (see Tables 2 and 4) between the intensities observed in the Iowa rocket flights and those computed through the vector addition of the main and transient field. As previously noted, the main field is extrapolated by means of the harmonic coefficients published by *Finch and Leaton* [1957]; the transient field, by means of a Taylor series. The small disparity between computed and observed equatorial results, about 1 part in 600, prompted a calculation of the variation with altitude of the location of the magnetic equator (where $Z = 0$) for the epoch 1955.0; this problem relates closely to the equatorial cosmic-ray studies of *Simpson and others* [1956; *Katz and others*, 1958]. The results for this epoch agree to within half a degree with those published for 1945.0 [*Zmuda*, 1958], using harmonic coefficients computed by *Vestine* [1947], and will not be reproduced here since the conclusions for the two epochs are identical: the cosmic-ray equator lies, in general, outside the region of space bounded by the dipole and the surface dip equator.

To account for the small differences of Table 4 for the SUI flights, several possibilities exist; but without additional flights and studies, it appears, the cause cannot be incontrovertibly determined. Nevertheless the results are compatible with the assumption that a time-independent equatorial ring current exists at a very

TABLE 4. Comparison between Observed and Computed Values of Scalar Intensity, for Points below Current System

All intensity values are in units of gammas.

Latitude, km	Observed Values				Observed Minus Computed Values*			
	SUI-87	SUI-86	SUI-83	Aerobee NRL-33	SUI-87	SUI-86	SUI-83	Aerobee NRL-33
40	33, 800	33, 367	32, 872	50, 984	41	67	-55	-461
60	33, 480	33, 036	32, 593	50, 470	42	53	-46	-454
80	33, 170	32, 717	32, 287	49, 975	49	47	-44	-435
95	32, 930				44			
96		32, 486				63		
104			31, 910				-59	
162				47, 980				-395

* With the exception of the computed Aerobee value, which is the value of the main field, the computed values are obtained through the vector addition of the main and transient field.

great distance above the earth's surface and with a direction of flow in general west to east. This direction of flow is opposite to that associated with the ring currents suggested as the cause of the main phase of a magnetic storm.) In the equatorial region the field of this ring current would be horizontal, directed northward like the main and total field, and practically constant with altitude as well as with horizontal displacement for points near the earth's surface as a consequence of these properties, the observed values would exceed by a constant amount those computed here, since the computed values do not contain any contributions from time-independent sources, such as ring currents, of external origin. This characteristic is approached with the results for SUI-86 and SUI-87, flights displaced from each other by about 50 miles. In a continuation of this reasoning, the expected constant difference for SUI-83 is a positive one; actually, the difference is negative, a finding that would ordinarily dispel further consideration of this hypothesis of a ring current. Anticipating a later result, however, negative (or westward) currents flow in the ionospheric region where SUI-83 was launched; and it may be that the field of this current (which could account for the negative differences found for this flight) predominates over the ring-current field in this region.

Although the existence of this time-independent eastward-flowing ring current cannot be considered established, it appears that the pos-

sibility nevertheless merits further study, with special attention to the effects of this field on the trajectories of cosmic rays.

There is only fair agreement between the computed and observed data for the White Sands flight, Aerobee NRL-33. In view of its sign and variation with altitude, the difference here may be attributed to the field of a localized anomalous structure, of origin internal to the earth, with an inverse magnetization and decreasing with altitude like $(a/r)^4$, where a is the radius of the earth. (If the ring current exists, its field at White Sands will be in large measure masked by this large local field.) The difference is not attributable to errors in the harmonic coefficients: the maximum decrease of any part of the main field is $(a/r)^8$. If vector data were available at one point, say at the earth's surface, additional characteristics of the anomalous field could be determined by means of relationships obtained like (8) for the transient field.

The White Sands negative anomaly (which as anomalies go is large in magnitude and, judging by the relatively slow decrease with altitude, probably large also in horizontal extent) may contribute to the formation of the region of low intensity of radiation lying between the two Van Allen radiation belts [Van Allen and Frank, 1959]. From White Sands with geomagnetic latitude 41°N , the line of force of the centered dipole field extends out to a maximum distance of about 2.3 earth radii from the center of the

TABLE 5. Values of Current Intensity

Rocket Flight	Current Intensity, amp/km		
	Integral of Computed Current Density	Computed from Surface Field Values Using Infinite-Sheet Approximation	Clockwise Angle between Direction of Current Flow and Direction of Geographic East, degrees
SUI-83	-4	124	202*
SUI-86	96	148	20
SUI-87	225	350	4
Aerobee A-11†	228		

* See text for discussion of the direction of current flow for SUI-83.

† Surface data were not available for the flight Aerobee A-11.

earth and passes through the region of low intensity. The field of the negative anomaly would permit particles originally trapped on this line of force to penetrate into the atmosphere deeper than they would in the absence of the negative anomaly. Deeper penetration will be accompanied by increased atmospheric absorption and scattering, and hence a reduction of intensity through removal of the particles from the trapped region. Since the charged particles oscillate along a line of force and drift (negative charges drift to the east; positive charges, to the west), a reduction in intensity for the White Sands area would contribute to a reduction in the entire volume lying between the two Van Allen belts.

The characteristics of the equatorial currents presented in Figure 3 and Table 5 show that the currents vary considerably with position and with local time. The variation with position at a fixed time is exemplified in the results of a comparison of the currents determined in the flights SUI-86 and 83.

These rockets were launched on successive days but at practically the same time of day, around 1400 local time, from positions displaced from each other by about 280 miles. The magnetograms for the days of launch were very similar and, as previously noted for SUI-86, showed for the time of launch a closeness of magnetic values for the three near-by observatories. At the three Scripps observatories, the launch-time similarity in magnetic elements and the differences in peak values suggest that the overhead currents detected in these two rocket flights are characteristic primarily of the currents

producing the normal daily magnetic variation, although a contribution from the electrojet currents cannot be completely discounted. The currents for SUI-86 are located in two shells, each about 12 km thick, with a direction of flow in each shell that makes a clockwise angle of 20° with the direction of geographic east, or of 101° with respect to the direction of the horizontal component of the total observed intensity. The currents for SUI-83 are also in two shells, but thinner ones; and for each shell the average of the absolute magnitude of the current density practically equals that of the corresponding average for the currents of SUI-86 (about 5 amp/km² for the lower and about 2 amp/km² for the upper shell). However, in addition to a difference in current intensity, stemming from a difference in thickness, the direction of flow in the lower shell for SUI-83 is considerably different from that for SUI-86, although the direction of the upper currents is practically identical, being 22° with respect to the direction of geographic east for SUI-83 and, as previously noted, 20° for SUI-86. In general, the direction of current flow here will be related to the direction of geographic east. The direction with respect to the magnetic elements is readily obtainable with the aid of the data listed in Table 1, where, for example, it is noted that for all the Iowa flights the declination of the total (time-independent and transient) intensity is about 9°E.

For SUI-83, the direction of flow of the current in the lower shell, a current listed as negative in Figure 3, is oppositely directed with respect to that of the current in the upper shell;

the negative current, makes a clockwise angle of 202° with the eastern direction. (With reference to the intensity curve for SUI-83 in Fig. 1, the presence of the negative current is indicated by the decrease, at an altitude of 104 km, of the slope of the intensity curve.) Also, the current intensity of the negative current (-18 amp/km) exceeds that of the positive current ($+14$ amp/km), so that the net current intensity has the direction of the negative current, making a clockwise angle of 202° with the direction of geographic east. The existence of the negative current intensity is not predictable using, with the transient field, sheet-current approximations and either surface values or values for points directly below the current sheet. In sheet-current approximations (where, as is the mathematical custom, the sheet is assumed to be of zero thickness), whether the sheet is infinite in horizontal extent or is spherical, the direction of flow of the current intensity is estimated from the direction of the transient field. For example, as previously noted, if for points below the current system the horizontal component of the transient field lies along the northern direction, the direction of flow of the current intensity is toward the east, in accord with the familiar right-hand rule connecting direction of current flow and magnetic field. At the time of this flight, for points below the currents the horizontal component of the transient field was directed principally northward and would be related in sheet-current approximations to a direction of current flow identical with that of the positive currents flowing in the upper current shell. It may be that, in the regions above that reached by this rocket, an additional positive current of sufficient density flows, so that the net current intensity may be positive, a result that would be compatible with that obtained with sheet-current approximations. An alternative possibility is that the negative current is *time-independent*. In this case, the positive current in the upper shell would relate to the transient field, and the negative current would connect with a steady field whose value for points below the current system would equal the negative difference between the computed and observed scalar intensities listed in Table 4.

It may be mentioned here that the magnitude of the current intensity obtained using a

sheet approximation and the surface value of the transient field is 124 amp/km. This value is much larger than that obtained by integrating the current density, -4 amp/km when negative currents are included and 14 amp/km when they are excluded. In general, for SUI-83 the values of the current intensity obtained from the observed field changes are practically in total disagreement with the values computed in sheet-current approximations.

The rockets SUI-87 and Aerobee A-11 were fired on different days but at times when the overhead currents are at or near their maximum; the currents now consist of the electrojet and the current that produces the peak normal daily magnetic variations. The currents here are considerably larger than the currents for SUI-83 and 86, and the current density for Aerobee A-11 is about twice the average in the lower shell penetrated by SUI-87. In SUI-87 a second current shell was discovered, separated from the lower shell by a region of no current about 10 km thick.

With the assumption that the magnitude of the electrojet current is the difference between the current for SUI-86 and SUI-87 (as appears likely, considering the shape of the curve of the daily magnetic variation), the current intensity of the electrojet is about 130 amp/km. With the Jarvis transient field as a criterion, the direction of current flow of the electrojet, a direction assumed to be that of the total current flow, makes a clockwise angle of 4° with the eastern direction and, for this area, is practically along the magnetic equator for 1955 as shown in the dip chart of the U. S. Navy Hydrographic Office.

The current intensities for SUI-86 and 87 (96 and 225 amp/km, respectively) differ from (1) the intensities computed using the approximation of an infinite current sheet, 148 and 350 amp/km, respectively; and (2) the maximum equatorial value obtained in harmonic analysis of the surface transient field, 50 amp/km, which occurs at about 1100 local time. For a detailed discussion of the results from harmonic analysis of the surface field, the reader is referred to a paper by Chapman [1951]. (The infinite-sheet currents of Table 5 are approximately halved if the transient intensity is treated as a departure from the daily mean value.)

The difference between the current intensity computed as the integral of the current density and that computed with sheet-current approximation and surface data may be caused by the following effects, in addition to the effect previously suggested for the difference in SUI-83. First, at altitudes above those discussed here, there may be currents with such characteristics that the present differences are eliminated. Second, the differences may be due to poloidal currents, which, as previously noted, cannot be estimated from field values obtained outside the current system, and hence are not included in the values of sheet currents estimated from surface data. If the poloidal currents exist, however, their horizontal components are contained, in addition to those of the toroidal currents, in the current densities computed from the rocket data. Thus the difference in the two sets of current intensities may be attributable to poloidal currents flowing in the regions traversed by the rockets.

Since it is not possible here to separate the transient field into its primary and secondary parts, the currents (including the sheet currents) computed here, where the transient field changes are related directly to the current densities, may require a maximum reduction of about 30 per cent, roughly the contribution, at the earth's surface, of the secondary field to the total transient field.

The magnitudes of the currents for the Iowa flights, shown in Figure 3, have similarities to as well as dissimilarities from the currents computed by Cahill [1959b]. For points inside the current region Cahill used a transient field that equaled the difference between the observed values and those computed using an inverse-cube-law decrease from a point located below the current region. Some of the dissimilarities appear to be due to varying judgments on the shape of the curve to be drawn through the observational points; others, of a more basic nature, relate to the method of analysis as well as to the interpretation of the data. A fundamental difference in the computed currents is the presence of the positive currents the author computes for SUI-83 for the region above 109 km. With reference to the scalar intensity curve for this flight (see Fig. 1), the author, using the previously treated characteristics of the

transient field, associates the increase in slope of the intensity curve for this region (a slope different from that for points below the current system) with a positive, or general west-to-east, current. Also, the directions of the current flow determined by the author are rotated in varying degrees with respect to those noted by Cahill, who discussed only the east-west component of current density, whereas the present author computed the north-south as well as east-west component. The difference here stems from the use in this report of the following related characteristics of the transient field: the X and Y components vary with altitude in a similar manner; and the current flow is normal to the direction of the transient field, not to the direction of the total, transient and steady, field.

Summary and conclusion. This paper has presented a method of analysis that, when applied to the available rocket data on the scalar magnetic intensity, yielded information on the upper-air vector field and the vector current density of the ionospheric currents that produce the daily magnetic variation. The method, applicable to data for all regions, nonequatorial as well as equatorial, may also be used in the study of upper-air observations on the vector magnetic intensity.

In the analysis it is necessary to treat first the observational data for points located between the earth's surface and the base of the current system. The measured values of scalar intensity are compared with values obtained through the vector addition of the main and transient field, extrapolated from the earth's surface. The agreement between the observed and computed values is very good for the equatorial flights (within about 60 γ) but only fair for the White Sands flight. The calculations for White Sands indicate the presence of a negative anomaly, which may contribute to the reduction of intensity observed in the region between the two Van Allen radiation belts. All the results, however, indicate that no large, if any, corrections are required for the harmonic coefficients of the main field.

For points inside the currents, the vector transient field, whose radial variation yields the vector current density, was separated from the observed scalar intensity through expressions derived with the aid of the boundary conditions

and the general form of the electromagnetic solutions. For the equatorial flights considered here, these expressions yielded current magnitudes practically identical with those obtained by using as the transient field the difference between the observed scalar intensity and the intensity obtained by continuing into the current region and, without a change in slope, the intensity variation found in the region below the current system. Thus, the latter method of obtaining currents (which cannot be used for nonequatorial regions, since there the transient field is no longer in the same general direction as the total field) suffices in cases where the equatorial currents constitute the sole interest, to the exclusion of, say, the also important consideration of the precision with which one can extrapolate the surface field upward. In both approaches for obtaining currents it is necessary to relate initially the direction of current flow to the direction of the transient field that existed at the time of the rocket flight, not, as is sometimes done, to the direction of the total (transient and time-independent) field.

The equatorial electrojet has an intensity of 30 amp/km and flows along the magnetic equator, making a clockwise angle of 4° with the geographic eastern direction. The other equatorial currents, which also flow in the E region of the ionosphere, form complicated patterns that vary considerably with position and time but nevertheless have the common characteristic of being confined to two shells separated by a region of little or no current. One of the current shells, the lower shell associated with UI-S3, contains a nearly westward current which is not predictable through the use of surface data and sheet-current approximations, including the results obtained in the customary harmonic analysis of the surface field of the daily magnetic variation. This unusual current should be investigated further. If there are no large eastward currents flowing in the region above that explored by the present rockets, it may be that the discovered westward currents are *time-independent* and, therefore, should not be related to those anticipated from surface studies of the daily magnetic variation.

In the interpretation of the graph of the daily magnetic variation, a problem exists in the

choice of the zero-reference value of the time-varying field. The two possibilities considered, the surface field used as a deviation from a mean and then from a minimum value, lead to different night-time characteristics. Thus the difficulty may be removed through analysis of night flights, preferably in the equatorial region where the expected differences are large.

Finally, the discussion on the general electromagnetic solutions associated with the unavoidable volume distributions of ionospheric currents may, in itself, have satisfied a need for a treatment of this problem, in relation both to present measurements on the scalar magnetic intensity and to future upper-air data on the vector magnetic intensity.

Acknowledgments. The author gratefully acknowledges the computational assistance of Martha Neuman and William Dickens, of the Applied Physics Laboratory of The Johns Hopkins University; and the cooperation of Drs. W. Cahill, Jr., and J. Heppner, who provided the experimental data used in the text.

REFERENCES

- Alexander, N., and C. Onwumechilli, Variation of the horizontal force near the magnetic equator, *Nature*, 180, 191-192, 1957.
- Baker, W. G., and D. F. Martyn, Conductivity of the ionosphere, *Nature*, 170, 1090-1092, 1952.
- Bullard, E., and H. Gellman, Homogeneous dynamos and terrestrial magnetism, *Phil. Trans. Roy. Soc. London, Ser. A*, 247, 213-277, 1954.
- Cahill, L. J., Jr., Magnetic exploration of the upper atmosphere, unpublished dissertation, State University of Iowa, 148 pp., 1959a.
- Cahill, L. J., Jr., Investigation of the equatorial electrojet by rocket magnetometer, *J. Geophys. Research*, 64, 489-503, 1959b.
- Chapman, S., and J. Bartels, *Geomagnetism*, Oxford University Press, London, 2 vols., 1049 pp., 1940.
- Chapman, S., The equatorial electrojet as detected from the abnormal electric current distribution above Huancayo, Peru, and elsewhere. *Arch. Meteorol. Geophys. u. Bioklimatol.*, 4, 368-374, 1951.
- Chapman, S., Rockets and the magnetic exploration of the ionosphere, in *Rocket Exploration of the Upper Atmosphere*, edited by R. L. Boyd and M. J. Seaton, Pergamon Press, London, pp. 292-305, 1954.
- Egedal, J. The magnetic diurnal variation of the horizontal force near the magnetic equator, *Terrestrial Magnetism and Atmospheric Elec.*, 52, 449-451, 1947.
- Elsasser, W., Induction effects in terrestrial magnetism, *Phys. Rev.*, 72, 821-833, 1947.

- Finch, H. F., and B. R. Leaton, The earth's main magnetic field—epoch 1955. 0, *M. N. Geophys. Suppl.*, 7, 314-317, 1957.
- Heppner, J. P., J. D. Stolarik, and L. H. Meredith, The earth's magnetic field above WSPG, New Mexico, from rocket measurements, *J. Geophys. Research*, 63, 277-288, 1958.
- Katz, L., P. Meyer, and J. Simpson, Further experiments concerning the geomagnetic field effective for cosmic rays, Proc. Varenna Conference, Spec. Supple., *Nuovo cimento*, 8, 277-282, 1958.
- Maple, E., W. A. Bowen, and S. F. Singer, Measurement of the earth's magnetic field at high altitudes at White Sands, N. M., *J. Geophys. Research*, 65, 115-126, 1950.
- McNish, A., Progress of research on magnetic diurnal-variations at the Department of Terrestrial Magnetism, Carnegie Institution of Washington, *Intern. Assoc. Terr. Magn. Elec. Bull.* 10, 271-280, 1937.
- Queenby, J. J., and W. R. Webber, Cosmic ray cut-off rigidities and the earth's magnetic field, *Phil. Mag.*, 4, 90-113, 1959.
- Rothwell, P., Cosmic rays in the earth's magnetic field, *Phil. Mag.*, 3, 961-970, 1958.
- Simpson, J. A., K. B. Fenton, J. Katzman, and D. C. Rose, Effective geomagnetic equator for cosmic rays, *Phys. Rev.*, 102, 1648-1653, 1956.
- Singer, S. F., E. Maple, and W. A. Bowen, Evidence for ionospheric currents from rocket experiments near the geomagnetic equator, *J. Geophys. Research*, 66, 265-281, 1951.
- Singer, S. F., E. Maple, and W. A. Bowen, Dynamo currents and conductivities in the earth's upper atmosphere, *Nature*, 170, 1093-1094, 1952.
- Van Allen, J. A., and L. A. Frank, Van Allen radiation belt, *Natl. Acad. Sci. IGY Bull.* 21, 3-6, 1959.
- Vestine, E. H., I. Lange, L. Laporte, and W. E. Scott, *The Geomagnetic Field, Its Description and Analysis*, Carnegie Inst. Wash. Publ. 580, 390 pp., 1947.
- Vestine, E. H., Exploring the atmosphere with a satellite-borne magnetometer, in *Scientific Uses of Earth Satellites*, edited by J. A. Van Allen, University of Michigan Press, Ann Arbor, 1956.
- Zmuda, A. J., Limiting form of Taylor series used in extrapolating components of the geomagnetic intensity, *Trans. Am. Geophys. Union*, 37, 9-12, 1956.
- Zmuda, A. J., A method for analyzing values of the scalar magnetic intensity, *J. Geophys. Research*, 63, 477-490, 1958.

(Manuscript received August 4, 1959; revised October 13, 1959.)

The Simultaneity of Sudden Commencements of Magnetic Storms

V. L. WILLIAMS

*Goddard Space Flight Center
National Aeronautics and Space Administration
Washington 25, D. C.*

Abstract. Rapid-run magnetograms from the U. S. Coast and Geodetic Survey (USC & GS) and from other observatories located near the magnetic equator were used to study sudden commencements that occurred during the period October 1957 to September 1958. Analysis of these magnetograms yielded three important results: (1) the sudden commencement always occurred first in high or middle latitudes; (2) Little America, Antarctica, registered the sudden commencement first or second 85 per cent of the time; and (3) the apparent propagation velocities of the sudden commencement around the magnetic equator had average values between 1145 and 2835 km/sec.

INTRODUCTION

The degree of simultaneity of the sudden commencement has been under discussion for at least 80 years. Investigations on this subject [Adams, 1881; Angenheister, 1913; Bauer, 1910, 1925; Chapman, 1918; Chapman and Bartels, 1940; Chree, 1910; Dessler, 1958; Ellis, 1892; Faris, 1910, 1911; Rodes, 1922; and Tanakadate, 1934] have reported propagation velocities ranging in value from 50 km/sec to the speed of light. This wide range of velocities, which resulted from both theoretical and experimental studies of sudden commencements, shows the disagreement that has been prevalent on this topic.

ORIGIN OF DATA

During IGY many magnetic observatories operated rapid-run magnetographs. The time scale of the rapid-run magnetograph, 4 mm/min, is 12 times better than that available on the standard magnetograms, which were used in all previous comprehensive experimental studies of the propagation velocity [but see Tanakadate, 1934, for limited rapid-run measurements]¹

The sudden commencements for this study

¹V. B. Gerard [1959] reported a study of the simultaneity of three sudden commencements occurring in 1957. Although the present analysis was conducted in ignorance of Gerard's work, the measurements of the sudden-commencement times for November 6, 1957, the only common date for both studies, agreed within a few seconds.

were selected by the criterion that three of the five stations—College, Sitka, Fredericksburg, Tucson, and Honolulu—must report a registration of the event in the *Journal of Geophysical Research*. The fifteen sudden commencements occurring between October 1957 and September 1958 fulfilling this condition are listed in Table 1. Rapid-run magnetograms were secured for all thirteen of the USC&GS magnetic observatories for the approximate GMT dates listed in Table 1; the locations of the observatories are given in Table 2. For additional data on longitudinal time differences a circular letter was sent to thirteen magnetic observatories located around the magnetic equator, requesting data on the sudden commencements listed in Table 1.

TABLE 1. Sudden-Commencement Dates

Date	GMT, approximate
06 Nov 1957	1821
11 Feb 1958	0126
16 Feb 1958	1641
14 Mar 1958	1212
31 May 1958	1652
07 Jun 1958	0046
14 Jun 1958	1828
28 Jun 1958	0714
08 Jul 1958	0748
21 Jul 1958	1636
17 Aug 1958	0623
22 Aug 1958	0226
24 Aug 1958	0140
03 Sep 1958	0842
25 Sep 1958	0408

TABLE 2. Location of USC & GS Stations

Station	Abbr.	Geographic Latitude	Geographic Longitude, E	Geomagnetic Latitude	Geomagnetic Longitude
Point Barrow	PB	N 71.3	203.3	68.6	241.2
College	CO	N 64.9	212.2	64.5	255.4
Big Delta	BD	N 64.0	214.3	64.3	259.3
Healy	HY	N 63.9	211.0	63.6	256.6
Sitka	SI	N 57.1	224.7	60.0	275.4
Fredricksburg	FR	N 38.2	282.6	49.6	349.8
Tucson	TU	N 32.2	249.2	40.4	312.2
Honolulu	HO	N 21.3	201.9	21.1	266.5
Guam	GU	N 13.6	144.9	4.0	212.9
Koror	KR	N 7.3	134.5	-3.2	203.4
Marie Byrd	BY	S 80.0	240.2	-70.6	336.0
Little America	LA	S 77.2	197.8	-74.0	312.0
Wilkes	WK	S 66.3	110.5	-77.8	179.1

The stations that sent data are given in Table 3 along with their locations.

ANALYSIS OF THE DATA

Measurements were made on the USC&GS magnetograms by means of a millimeter scale and a magnifying glass. The first indication of deviation from a previous course was taken as the start of the sudden commencement for both the SC and SC*. This method, a suggestion of la Cour, helps eliminate instrument differences that occur at the different stations. Analysis was made only of the *H* component. Although a measuring accuracy of ± 1.5 seconds would not be too difficult to maintain, an inaccuracy arises from not being able to pick the same starting point for the commencement at the various observatories. This inaccuracy varies directly with the sharpness of the sudden commencement. For a disturbance of average sharpness the error is believed to be of the order of ± 5 seconds. The data were corrected for paper shrinkage, parallax, and time error, these corrections being in general small in comparison with the above reading error.

The USC&GS stations will be discussed first. In graphs A through O can be seen plots of the order of occurrence of the sudden commencements at the U.S.C. & G.S. observatories. (From this point on, the sudden-commencement date will be referred to by the letter designation appearing on the graph above the date.) Below the date is a small plot of the order of occurrence against time. In the center of the graph

the progression of the sudden commencement is traced by means of straight lines drawn from observatory to observatory. When the sudden commencement occurred at two stations at the same time, both stations were given the same order of occurrence number. An example of this situation can be seen in graph D for observatory 8. The local time is given across the top of the graph.

The tendency for sudden commencements to occur first at high latitudes is apparent in graphs A through O. To show this important feature more clearly, the order of occurrence was divided into three groups corresponding to early (1-4), middle (5-8), and late (9-12) arrivals. The percentage of occurrences falling in each group was then calculated for each station by dividing the number in each group by the total number of registrations at that station. Results are shown in Table 4. It is apparent that early registrations were most frequent at high latitudes, whereas the highest percentages in low latitudes fall in the middle category. The exceptionally high percentage of early arrivals at Little America is striking. Even more striking is the fact that nine of the sudden commencements occurred first at Little America.

The equatorial stations listed in Table 3 plus Koror and Guam will now be discussed. These sudden-commencement times are given in Table 5 in minutes and seconds. Even though the measurements were made by different observers the times are very close.

TABLE 3. Location of Additional Stations

Station	Country	Geographic Latitude	Geographic Longitude, E	Geomagnetic Latitude	Geomagnetic Longitude
Parmaribo (PA)	Surinam	N 5.8	304.8	17.0	14.5
Bangui (BA)	Fr. Eq. Afr.	N 4.6	18.6	5.0	88.6
Leopoldville (BI)	Bel. Congo	S 4.4	15.2	-3.1	83.5
Hollandia (HN)	Dutch N. Guiana	S 2.6	140.5	-12.5	210.3

The apparent propagation velocities for the sudden commencements were calculated for both the short and long great-circle distances between observatories. These velocities were calculated for the individual time differences and average time differences for the distances between Parmaribo and Hollandia, Koror and Leopoldville, Koror and Parmaribo. (Guam was not used, because its times and location were essentially the same as those of Koror. Bangui was not used because of the large amount of missing data.) The velocity calculations are given in Table 6.

Although there is some consistency in the apparent velocity between the combinations of stations for a given sudden commencement, some large differences appear. They probably result more from the errors in measurement than from actual differences over equatorial paths.

It was determined [Williams, 1959] that for the sharpest sudden commencements the time

between the first and last registration was smallest. Thus it appears that the more accurately the sudden commencement can be measured the faster will be the apparent propagation velocity of the commencement until the upper limit of the velocity is reached.

With these qualifications it is highly probable that actual equatorial velocities are in general greater than those given by the 'velocity from the average time difference' of Table 6.

CONCLUSION AND DISCUSSION

The results of this study can be summarized as follows:

1. The sudden commencements registered in the high and middle latitudes first.
2. The sudden commencement registered at Little America first or second about 85 per cent of the time.
3. The apparent propagation velocities of the sudden commencement around the magnetic equator had average values between 1145 and 2835 km/sec.

Some of the consequences of the above results are:

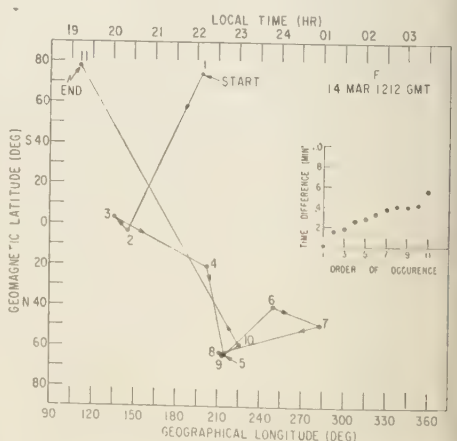
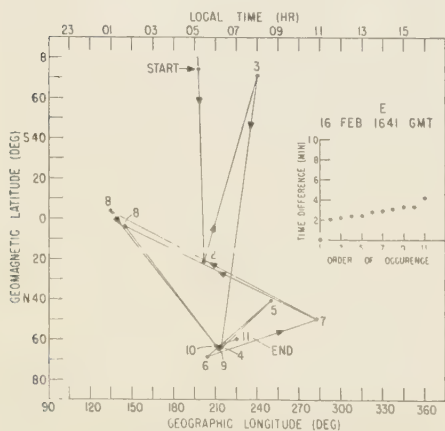
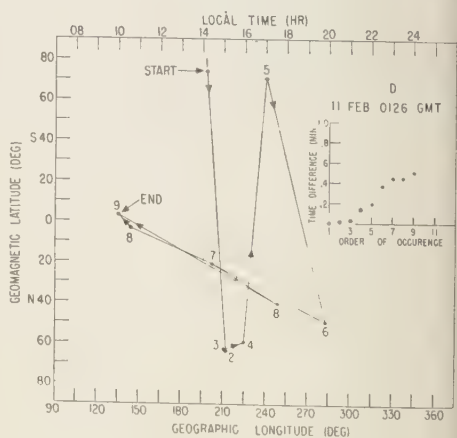
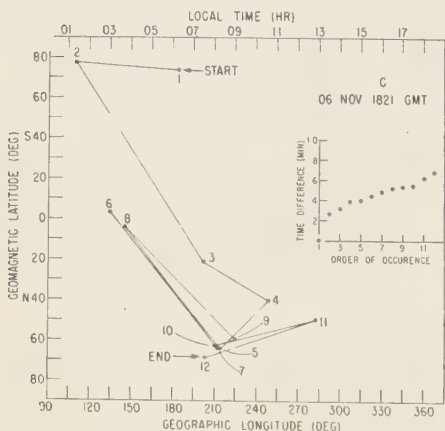
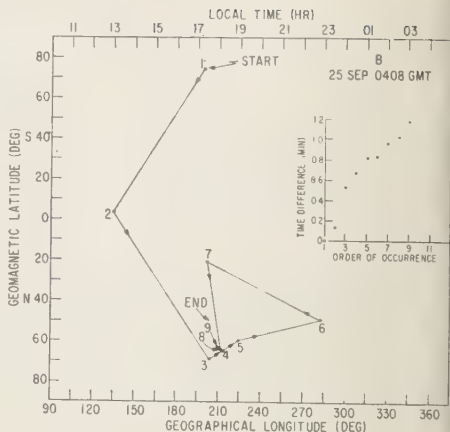
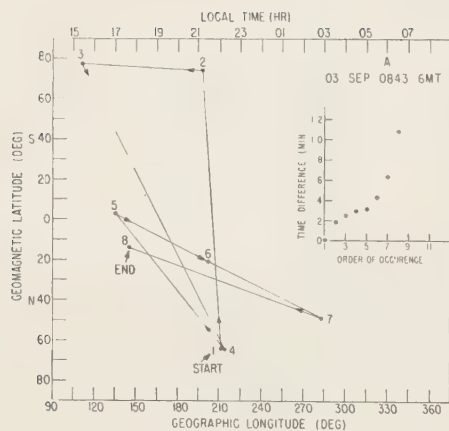
1. The high-latitude beginning lends support to Singer's [1957] model of a magnetic storm.
2. The high-latitude beginning is consistent with Heppner's [1955] suggestion that the sudden commencements begin in the ionosphere of the auroral zone.

3. The propagation velocity of 130 km/sec calculated by Dessler [1958] for a sudden commencement propagated around the magnetic equator is not in agreement with the velocities resulting from this study.²

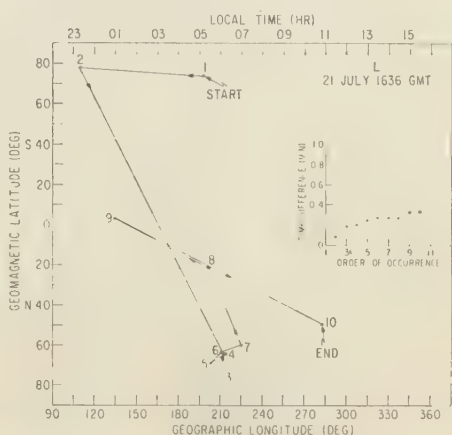
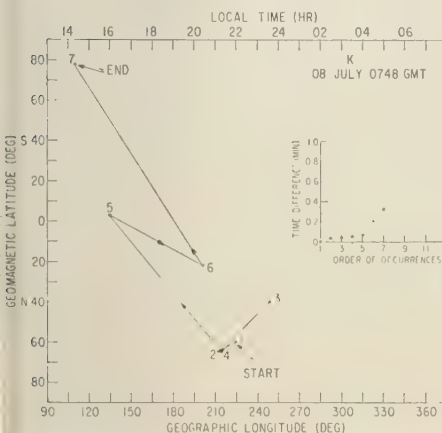
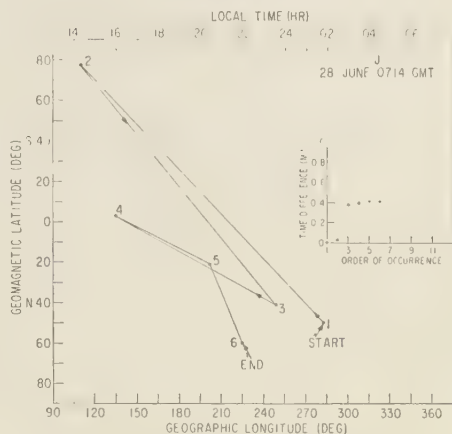
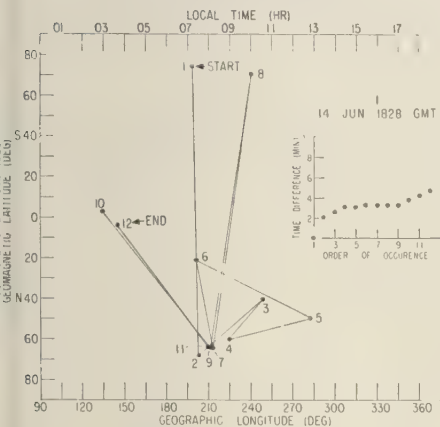
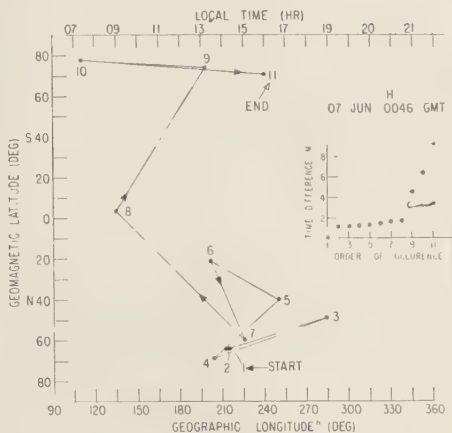
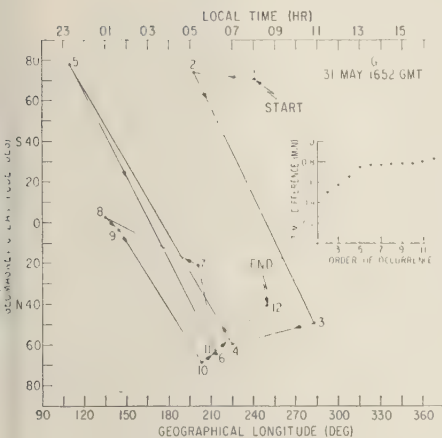
² Subsequent to the writings of this paper Green, Francis, and Dessler [1959] have published a new sudden-commencement propagation velocity. They give time differences that correspond to an apparent propagation velocity of approximately 1800 km/sec.

TABLE 4. 'Order of Occurrence'
Percentage for USC & GS Stations

Station	Order of Occurrence			
	1-4	5-8	9-12	Geomagnetic Latitude
B	50	33	16	68.6
O	46	46	38	64.5
D	54	46	0	64.3
Y	22	44	33	63.6
I	32	30	38	60.0
R	36	43	21	49.6
U	42	42	17	40.4
O	13	80	13	21.1
U	20	60	20	4.0
R	27	53	20	-3.2
Y	29	43	29	-70.6
A	85	0	15	-74.0
/K	54	18	27	-77.8



Order of occurrence for the sudden commencements at the U. S. Coast and Geodetic Survey magnetic observatories.



Order of occurrence for the sudden commencements at the U. S. Coast and Geodetic Survey magnetic observatories.

TABLE 5. Sudden-Commencement Time for Equatorial Stations

Sudden Commencement	PA		HN		KR		GU		BI		BA	
	min	sec	min	sec	min	sec	min	sec	min	sec	min	sec
A	42	40	x*	x	42	05	x	x	42	27	x	x
B	07	50	x	x	07	05	x	x	07	35	x	x
C	x	x	x	x	20	43	20	48	x	x	x	x
D	x	x	x	x	25	38	25	35	25	46	x	x
E	x	x	41	17	41	18	41	18	x	x	x	x
F	12	20	x	x	11	52	11	50	11	54	11	53
G	52	30	52	12	52	10	52	10	52	26	52	20
H	45	58	45	54	45	53	x	x	45	53	x	x
I	28	15	28	01	27	57	28	02	27	56	x	x
J	13	04	13	36	13	17	x	x	13	01	13	47
K	x	x	x	x	48	03	x	x	x	x	48	20
L	36	37	36	33	36	35	x	x	36	24	36	37
M	22	36	22	26	22	29	22	26	22	32	x	x
N	27	16	28	01	27	33	27	49	27	25	x	x
O	39	58	40	04	40	03	40	03	40	06	x	x

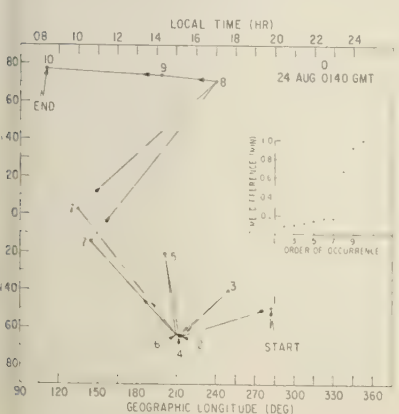
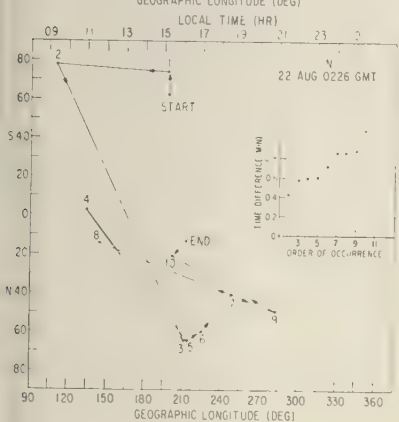
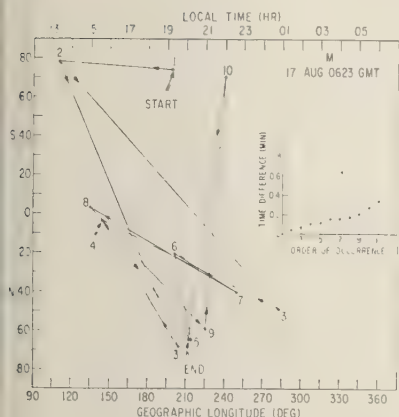
* x indicates missing data.

TABLE 6. Apparent Sudden Commencement Propagation Velocity (Expressed in kilometers per second)

Sudden Commencement	PA to HN		KR to BI		KR to PA	
	Long	Short	Long	Short	Long	Short
A	x*	x	1220	600	590	550
B	x	x	860	430	460	430
C	x	x	x	x	x	x
D	x	x	3340	1660	x	x
E	x	x	x	x	x	x
F	x	x	13370	6630	740	690
G	1210	1010	1670	830	1040	960
H	5440	4560	∞ †	∞ †	4160	3840
I	1550	1300	26740	13260	1150	1070
J	680	570	1670	830	1600	1480
K	x	x	x	x	x	x
L	5440	4565	2430	1210	10040	9610
M	2170	1830	8910	4420	2970	2750
N	480	410	3340	1660	1220	1130
O	3620	3040	8910	4420	4160	3840
Average time difference with probable error, sec	16.6 \pm 5.2		10.1 \pm 2.6		17.7 \pm 4.1	
Great-circle distance, km	21740	18260	26740	13260	20780	19220
Velocity from average time difference, km/sec	1450 \pm 450	1220 \pm 380	2835 \pm 735	1405 \pm 365	1240 \pm 290	1145 \pm 260

* x indicates missing data.

† Infinite velocity means that the registration time was the same at both stations within the accuracy measurement.



order of occurrence for the sudden commencements at the U. S. Coast and Geodetic Survey magnetic observatories.

Acknowledgements. The author gratefully wishes to acknowledge many valuable discussions with Dr. J. P. Heppner and T. L. Skillman during the course of this work.

An expression of appreciation is extended to Dr. J. Veldkamp, Royal Netherlands Meteorological Institute (Hollandia and Paramaribo); Dr. P. Herrinck, Service Météorologique Belgium Congo (Léopoldville); and R. Godiver, Office de la Recherche Scientifique et Technique Outre-Mer (Bangui). The sudden-commencement data these gentlemen sent from their respective equatorial stations enabled the author to make calculations for the propagation velocity of the sudden commencement around the magnetic equator.

REFERENCES

- Adams, W. G., Comparison of declination magnetographs at Kew, Stonyhurst, Coimbra, Lisbon, Vienna and St. Petersburg, *Brit. Assoc. Rept.*, pp. 201-209, 1880; On magnetic disturbances and earth-current, *Brit. Assoc. Rept.*, pp. 463-474, 1881.
- Angenheister, G., Propagation velocities of magnetic disturbances and pulsations, *Nachr. Ges. Wiss. Göttingen, Math.-Physik. Kl.*, 4, 565-581, 1913.
- Bauer, L. A., Beginning and propagation of the magnetic disturbance of May 8, 1902, and of some other magnetic storms, *Terrestrial Magnetism and Atmospheric Elec.*, 15, 9-20, 1910.
- Bauer, L. A., and Peters, W. J., Regarding abruptly-beginning magnetic disturbances, *Terrestrial Magnetism*, 30, 45-68, 1925.
- Chapman, S., On the times of sudden commencements of magnetic storms, *Proc. Phys. Soc. (London)*, 30, 205-214, 1918.
- Chapman, S., and J. Bartels, *Geomagnetism*, Oxford University Press, vols. I and II, 1049 pp., 1940.
- Chree, C., Time measurement of magnetic disturbances, *Proc. Phys. Soc. (London)*, 26, 137-153, 1914; also 23, 49, 1910.
- Dessler, A. J., The propagation velocity of worldwide sudden commencements of magnetic storms, *J. Geophys. Research*, 63, 405-408, 1958.
- Ellis, W., On the simultaneity of magnetic variations at different places on occasions of magnetic disturbance, and the relations between magnetic and earth current phenomena, *Proc. Roy. Soc. (London)*, 52, 191, 1892.
- Faris, R. L., Times of abruptly beginning magnetic disturbances, as recorded at the Coast and Geodetic Survey Magnetic Observatories, *Terrestrial Magnetism and Atmospheric Elec.*, 15, 93-105, 1910.
- Faris, R. L., Data for abruptly beginning magnetic disturbances 1906-1909, No. 1 *Terrestrial Mag-*

- netism and Atmospheric Elec.*, 16, 385-108, 1911; also No. 11, 16, 163-204, 1911.
- Gallet, R. M., The very low-frequency emissions generated in the earth's exosphere, *Proc. IRE*, 47, 211-231, 1959.
- Gerard, V. B., The propagation of world-wide sudden commencements of magnetic storms, *J. Geophys. Research*, 64, 593-596, 1959.
- Green, M. I., W. E. Francis, and A. J. Dessler, Refraction of hydromagnetic waves in the geomagnetic field, *Bull. Am. Phys. Soc.*, 4, 360, 1959.
- Heppner, J. P., Note on the occurrence of world-wide S.S.C.'S during the onset of negative bays at College, Alaska, *J. Geophys. Research*, 60, 29-32, 1955.
- Rodes, L., On the non-simultaneity of magnetic storms, *Terrestrial Magnetism and Atmospheric Elec.*, 27, 161-166, 1922.
- Singer, S. F., A new model of magnetic storms and aurorae, *Trans. Am. Geophys. Union*, 38, 175-191, 1957.
- Tanakadate, C. R., Assemblee de Lisbonne, September 17-25, 1933, *Union géodésy géophys. internationale assoc. magnetism Elec. Bull.* 9, 149-157, 1934.
- Williams, V. L., The simultaneity of sudden commencements of magnetic storms, unpublished master's thesis, University of Maryland, 1959.

(Manuscript received September 11, 1959;
revised October 13, 1959.)

Geomagnetic Storm Theory

J. H. PIDDINGTON

*Commonwealth Scientific and Industrial Research Organization
Radiophysics Laboratory
Sydney, Australia*

Abstract. A discussion of the two-gas theory of the transmission of geomagnetic disturbances through the atmosphere (to several earth radii) is extended, with the following results: (i) The central problem concerning the main phase of a geomagnetic storm is the mechanism of penetration of solar ions into the geomagnetic field. An explanation is given depending on a combination of a uniform electric space-charge field and a system of irregular fields. (ii) A model of the main phase of a geomagnetic storm is given, the principal feature of which is a 'magnetic tail' extending from the earth on the dark side. (iii) The model may help to explain some other effects: the Gegenschein, electrons with auroral energies, the location of the Van Allen zones, and diurnal cosmic-ray variations. (iv) All observed geomagnetic disturbances have their sources *initially* in current systems in the lower ionosphere. Some are subsequently maintained by current systems in the earth itself and in the region of interaction between the solar and terrestrial plasmas. Others, mainly polar and equatorial, are maintained by ionospheric currents driven by space-charge electric fields. (v) Any ring current outside the geomagnetic field could cause an increase in the horizontal component. A westward-flowing ring current embedded in the field could cause either an increase or a decrease in the horizontal component. The basic effect is not the current but a sustained inward or outward mechanical force on the material in which the current flows.

1. INTRODUCTION

It is widely agreed that geomagnetic storms and associated effects result from the arrival at the earth of a cloud of neutral, ionized gas moving out from the sun. In the early development of this theory [Chapman and Ferraro, 1931] it was shown that, as the cloud approached the earth, electric currents should be induced on its surface, with notable geomagnetic effects.

It is now known that the space between the solar-gas cloud and the earth is filled with electrically conducting gas whose effects, through self-induction, must be considered. In an earlier paper [Piddington, 1959] the problem of the transmission of geomagnetic disturbances through this medium was considered. Because of their immediate significance in magnetic storm theory these results are briefly repeated in section 2.

Two conclusions affecting magnetic storm and auroral theory follow from these results:

(a) Geomagnetic storm theory at present is unable to explain any significant penetration of the solar cloud into the geomagnetic field. Since there appears no possibility of explaining the main phase of a storm without substantial penetration, this becomes a central problem.

(b) The basic, if not sole, effect needed to provide the main phase is an outward mechanical force on the gas at a distance of a few earth radii.¹ With this conclusion in mind an attempt is made in section 3 to clarify some earlier discussions about the possible effectiveness of ring currents in explaining the main phase.

The main objectives of the present paper are, first, to describe a mechanism which may provide substantial penetration of solar ions into the geomagnetic field,² and, second, to outline a new theory of the main phase of a geomagnetic storm. The principal feature of this theory is that magnetic lines of force which, before the storm, extended to several earth radii become frozen into the solar wind and are carried away

¹ As a result of this force the series of events generally referred to as a 'hydromagnetic phenomenon' follows: the gas moves in the magnetic field; electric fields and currents are induced; the gas experiences a force of electromagnetic origin which opposes the original force; and, finally, the magnetic field is modified. However, all these are secondary effects which must follow the application of the original force.

² Apparent objections to mechanisms already advanced are given.

to form a 'magnetic tail,' which extends to large distances on the dark side of the earth.

Apart from geomagnetic effects, other phenomena may perhaps be explained in terms of the storm theory. These are the Gegenschein or counter glow, the existence of electrons of near-auroral energies, the location of one or even both of the Van Allen radiation zones, and the diurnal variation of cosmic-ray intensity.

2. TRANSMISSION OF GEOMAGNETIC DISTURBANCES TO THE EARTH'S SURFACE

It is now known that the whole of the space occupied by the geomagnetic field is filled with ionized gas. Thus, to relate the observed geomagnetic fluctuations to the original disturbance, the transmission of electromagnetic waves through a magnetoionic medium must be studied.

For waves of frequency much lower than any characterizing the medium (ion gyrofrequencies and relevant collision frequencies), transmission is by hydromagnetic waves in a single fluid [see, for example, *Piddington*, 1955], and the transmission characteristics depend mainly on the mass density and 'effective' electrical conductivity of the medium. A formula for this conductivity, σ_3 , has been given by *Cowling* [1956, equation 30]. Its value depends critically on the proportion of neutral atoms present, and at distances of a few earth radii, where the electron density is $\sim 10^3 \text{ cm}^{-3}$, this proportion is likely to lie between 10^{-6} and 10^{-3} [*Shklovskii*, 1958].³ The corresponding values of σ_3 are 10^{-15} emu, or less, for a magnetic field of 10^{-3} gauss. Such a low value of σ_3 would permit direct invasion by an external magnetic field (by a process of diffusion) in a matter of minutes. This would mean that the presence of the ionized gas was not particularly significant, because of its low electrical conductivity.

It may be shown, however, that a more general approach to the transmission problem is required; this is provided by a study of the separate equations of motion of the ion-electron plasma and the neutral atom gas [*Piddington*, 1956]. The former moves to some extent independently

under the influence of the electromagnetic field. For waves of sufficiently low frequency the two approaches yield the same value of σ_3 in ternary gas [*Piddington*, 1957, equation 8].

In the earth's atmosphere this 'two-gas' theory is required⁴ and provides useful information about the transmission of geomagnetic disturbances [*Piddington*, 1959]. The application of the theory shows that the waves do not travel as simple hydromagnetic waves but that the medium is dispersive.

For the purpose of developing geomagnetic storm theory it is helpful to divide the whole atmosphere into two regions, an upper (U) region lying above a few hundred kilometers and a lower (L) region. Transmission in the U region is effectively by hydromagnetic waves in the ion-electron plasma *alone* into which the magnetic lines of force are effectively frozen because of the high value of σ_3 for the plasma *alone* ($\sim 10^{-8}$ mu). The field and plasma oscillate through the neutral atom gas without appreciably disturbing the gas (in spite of the fact that it reduces σ_3 by a factor of 10^8 or more).

In the L region the neutral atoms render the medium truly dispersive and transmission is similar to that through a solid conductor, though it is by *diffusion* and not by hydromagnetic action.

These results are helpful in understanding geomagnetic storm phenomena and are first discussed briefly in connection with sudden commencement (SC) storms. This effect is caused by a sharply bounded cloud of solar plasma impinging on the outermost part of the earth's magnetic field [*Chapman and Ferraro*, 1940]. A current of density \mathbf{j} is induced, causing a *local* distortion of the field \mathbf{H} , given by

$$\text{curl } \mathbf{H} = 4\pi \mathbf{j}$$

This current system cannot directly affect the field at the earth since the U region is an excellent shield. However, the U region may transmit different types of hydromagnetic wave as does the L region. These waves result in a movement of the lower boundary of the U region: at low latitudes this is pushed bodily into the interior of the L region; in high latitudes there is a sideways component of motion. The entry of the U region into the L region causes a cc

³ These values are much greater than that of *Parker* [1958a], who assumed a radiation field corresponding to full black-body radiation at 10^4 K in the regions concerned.

⁴ Except for waves of period $\gg 5$ hours in the U region and $\gg 4$ months at an earth radius or so

sion of the field there, and this diffuses through the L region and is recorded at the earth's surface.

This reasoning led to the conclusion that the present systems *directly* responsible for all observed geomagnetic disturbances flow in the F region [Piddington, 1959]. This region extends from the F region of the ionosphere to below the earth's surface. The observed disturbances are initiated by currents in the E region caused by the diffusion of magnetic lines of force through the E region. As diffusion continues, the current system is transferred to the earth itself, the transfer being complete in a period of a few minutes. Except in particular cases mentioned above, currents cannot continue to flow in the ionosphere, because of its limited electrical conductivity.

The quasi-stable situation now attained comprises two current systems: one inside the earth, which is *directly* responsible for observed magnetic changes, and one at the interface of the solar stream and geomagnetic field. The latter is *directly* responsible for observed disturbances, and, if it varies, a corresponding geomagnetic disturbance will be observed some minutes later. Only in these two regions are the two requirements for quasi-steady current systems satisfied: high electrical conductivity and a mechanical force to balance the electromagnetic force $\mathbf{j} \times \mathbf{H}$. In the lower ionosphere the force is negligible, owing to the large inertia of the neutral gases; the conductivity, however, is too small. In the U region the plasma has sufficient conductivity but cannot provide a mechanical force except during the brief period taken for a geomagnetic wave to pass.

The above discussion refers particularly to storm-time variations whose equivalent current systems are essentially symmetrical about the earth's axis. The initial phase, referred to in the following sections as the SC, is essentially a world-wide increase in the horizontal component of the geomagnetic field. The main phase, which follows an hour or so later, is a general decrease in the horizontal component. It appears that all SC's are *initiated* by E -region currents, *directly maintained* by earth currents, and *indirectly supported* by currents at several other radii. The over-all picture is a hydromagnetic reformulation of that of Chapman and Ferraro.

As will be shown below, current systems in these three regions also account for the main phase, but the distant current system proposed (responsible for the 'magnetic tail') differs completely from the Chapman-Ferraro ring current.

When the storm-time variations are subtracted from the complete system, the longitude-dependent system of 'polar magnetic storms' remains. These are mainly due to accumulations of electric space charge in the ionosphere at high latitudes. The resulting electric fields, unlike the induction fields mentioned above, are not transitory but may maintain quasi-steady current systems *in the ionosphere*. The origin of the electric space-charge will be discussed in a later paper.

3. RING CURRENTS AND THE MAIN PHASE

If the SC increase results from an inward force at the top of the U region, the main-phase decrease might be expected to require a sufficiently powerful *outward* force applied to some part of the U region. Such a force is provided in the ring-current theory of Chapman and Ferraro [1933] and Martyn [1951] but not in other ring-current theories. Although this theory seems to meet great difficulties and an alternative is outlined below, it is nevertheless instructive to study it further with the new factor of highly conducting plasma in the U region.

In two recent papers, Hines and Storey [1958] and Parker [1958a] have expressed opposing opinions on ring-current theory. The former have confirmed that a slab of plasma with a dominating magnetic field cannot shield a magnetic disturbance, however high the electric conductivity of the slab; the disturbance is transmitted as a hydromagnetic wave as demonstrated elsewhere [see, for example, Piddington, 1955]. They conclude from this that 'a magnetic field would penetrate a highly conducting fluid just as it would penetrate a dielectric.'

Even when limited to the case of a weak perturbation of a dominating field, as was intended (Hines, personal communication), this conclusion might be queried on the ground that a particular magnetic line of force (identified by the particular ions which spiral around it) does not *penetrate* or diffuse into the surrounding plasma but pushes that plasma ahead of it as in the case of a sound wave. If the conclusion were accepted as general, it would follow that

a (westward) ring current situated quite outside the geomagnetic field could account for the main-phase decrease. In fact, any current outside the field could cause a decrease only in a period more than months [Parker, 1958a]. The relevant theory is not the weak perturbation theory of Hines and Storey but the neutral line theory of Sweet [1958] and others.

The important point to remember is that there is a great difference between the propagation of a hydromagnetic disturbance and the penetration or diffusion of a magnetic field into a plasma. Only in the latter case (which cannot occur appreciably in the outer atmosphere during the period of a storm) can lines of force be 'cut.' This conclusion is most significant in connection with the free flow of 'horns' of solar gas into the auroral zones as envisaged in the Chapman-Ferraro theory and as discussed in Piddington [1959] and section 4 below.

When Hines and Parker [1958] compared their views further they chose a more interesting situation, with the current path *inside* the undisturbed field. In this case, as will be shown, either of their physical pictures (see their Figs. 2 and 3, may be correct. The problem has not been fully formulated until the nature of the electric field responsible for the ring current has been specified.

First let us assume that the electric field is a potential field caused by connecting a battery to an insulated metal ring which had been placed in the undisturbed dipole field. As the current grows a magnetic field develops around the conductor which pushes back the perfectly conducting plasma and its frozen-in field. The situation is similar to that in a concentric cable except that the magnetic pressure in the space between the conductors may expand the outer conductor. This situation could only lead to an increase in magnetic field strength inside the wire loop.⁵

In the second formulation the electric field is an induction field caused by increasing the radius of the metallic ring in the dipole field. As the wire starts to cross the magnetic field, a westward (in the geomagnetic case) electric field and current are induced. The current flows only in

the thin outer surface of the wire into which the magnetic field has penetrated. The magnetic field due to this current is integral with the dipole field, corresponding to an outward movement of the lines of force of the field. The movement of the ring is opposed by an electromotive force of density $\mathbf{j} \times \mathbf{H}$, which, as previously concluded, is the fundamental physical requirement for the main phase of a storm.

As the lines of force outside the ring are pushed out, the space gradient in \mathbf{H} just inside the ring is increased and a rarefaction hydromagnetic disturbance propagates down through the U region. This causes the westward region to be lifted away from the L region whose field then proceeds to diffuse upward causing a general decrease in geomagnetic strength. The sequence of events is just the reverse of the SC increase. Thus a westward flowing ring current embedded in the earth's field may, if associated with a suitable outward mechanical force on the conductor, account for the main-phase decrease.

4. EXCLUSION OF SOLAR GAS FROM THE GEOMAGNETIC FIELD

From the above discussion and that of Parker [1958a] it is concluded that a ring current flowing outside the geomagnetic field could never account for the main-phase decrease. Furthermore, an essential feature of a ring current flowing inside the field is an outward mechanical force. In fact, if this outward force is exerted, all other effects follow automatically. In the case of the Chapman-Ferraro ring current the force is centrifugal provided by protons and electrons circling the earth in the same direction at high speeds.

If the solar gas were excluded altogether from the geomagnetic field and its plasma, it would clearly be incapable of providing any outward mechanical force. Substantial penetration of the field was possible according to the Sweet theory, but when account is taken of the plasma that fills the U region, penetration no longer appears possible⁶ [Piddington, 1959]. This rule

⁵ Except, of course, at the 'neutral plane' between the two separate field systems [Sweet, 1958].

⁶ In the absence of the plasma the lines of force of the geomagnetic field are 'cut' and cease to form complete loops, thus allowing 'horns' of gas to penetrate. When plasma is present this effect is prevented by self-induction effects; the lines of force are then merely flattened.

vides a difficulty which must be overcome. If any satisfactory theory, ring-current or otherwise, of the main phase can be developed. The magnitude of the problem of penetration can be seen by reference to a one-dimensional model of the solar gas being deflected by the earth's field [Ferraro, 1952]. The ions penetrate the field only a short distance before being reflected out again. The degree of penetration of the field into the gas cloud might be expressed as magnetic flux ψ , being the total flux that crosses the geomagnetic equatorial plane inside the gas cloud and within an angle (or range of latitude) of 1 radian. This flux may then be related to the colatitude θ , such that all the lines of force that cut the solar cloud cross the earth's surface at higher geomagnetic latitudes:

$$\psi = 1.3 \times 10^{17} \sin^2 \theta \text{ gauss cm}^2 \quad (1)$$

In Table 1 the values of ψ and θ are given for three conditions: the first assumes that penetration is to the auroral zone ($\theta = 23^\circ$); the second specifies a value of ψ which seems necessary to account for a major magnetic storm according to the drawn-out field theory [Piddington, 1959]; the third is the penetration, according to Ferraro's theory, of a proton-electron cloud of velocity 10^8 cm sec^{-1} (the penetration does not depend on the density of the cloud).

The value of ψ provided by the simple reflection process means that, of the total magnetic flux around the earth, less than 1 part in 2.6×10^6 is linked with the solar gas. This flux may be regarded as a thin magnetic sheet isolating the solar particles from the main part of the earth's field. Until this sheet is penetrated or blown away by the solar gas the gas could exert only an inward mechanical force, not an outward force as required. Even the removal of the sheet could only provide a very small decrease in the strength at the earth's surface, and perhaps many as 10^4 to 10^5 sheets would have to be removed to provide a major storm. Since the

main phase develops fully in a few hours, the time for the penetration or removal of each sheet would be of the order of a second.

In Ferraro's analysis collisions are neglected, and if an ion entering the sheet collided with one leaving they might be scattered into the magnetic field. The collision cross sections involved are so small, however, that this effect is not important.

As an alternative Parker [1958c] has suggested that collisions take place between blobs of space charge which form spontaneously in the inherently unstable medium where solar gas moving toward the earth interpenetrates gas that has suffered reflection. The inhibition of interpenetration of clouds of neutral ionized gas has been discussed by Kahn [1957]. According to this theory the two interpenetrating electron streams are first brought to rest, their kinetic energy of streaming being converted into plasma oscillations. Subsequently each of the heavy-ion streams interacts with the stationary electron gas and is also brought to rest. The net result is a violent frictional effect preventing any considerable interpenetration of gas clouds. Rather similar results are reached by Buneman [1958] in a discussion of a single electron stream and a single heavy-ion stream (an electric current).

Such an effect would be very important in magnetic storm theory: first, it would prevent the passage of solar gas through the interplanetary gas (the interplanetary gas would have to be pushed out ahead of the solar gas, forming a shock wave); next, it would allow scattering of the incident gas cloud into the earth's field; and finally, as Parker has suggested, it could provide electrons of near-auroral energies.

However, there appears to be an objection to the theory. The double electron stream will undoubtedly be unstable and will form small clouds of space charge at the expense of the streaming energy of the electrons. Because of their much greater mass the heavy ions will, up to this point, be slowed very little. The next step is for the electrons, having lost their streaming velocity, to be scattered by the small-scale fields. By energy-conservation arguments they may attain thermal velocities comparable with their original streaming velocity. At this stage all instability is removed, because of the random motions of the electrons. Subsequent inter-

TABLE 1. Penetration of Solar Gas into the Earth's Field

Condition assumed	$\theta, ^\circ$	ψ , gauss cm ²
Penetration to auroral zone	23	2×10^{16}
For magnetic storm	40	5×10^{16}
Simple reflection (Ferraro)	0.11	5×10^{11}

penetration of the heavy-ion streams is unimpeded. If this ion streaming tends to provide an electric current system, it is opposed by self-inductance effects and the electron gas is drawn along by the ion current to neutralize the current. The net result anticipated is a slight heating of the electron gas. In a solar stream interpenetrating stationary gas with velocity 10^8 cm sec⁻¹ the electron temperature must rise to $\sim 2 \times 10^4$ K.

The above argument is given in more detail in the appendix.⁷ The conclusion is that simple double-streaming cannot provide penetration or electrons of near-auroral energies.

5. ACCELERATION OF SOLAR ELECTRONS IN THE EARTH'S MAGNETIC FIELD

It would seem from the above discussion that, during their passage through the interplanetary gas, electrons in a solar cloud are heated sufficiently to eliminate plasma instability. On reaching the vicinity of the earth they and the heavy ions accompanying them would then be bounced off the earth's magnetic field without being scattered and without being able to penetrate the field and so cause the main phase of a magnetic storm.

Another space-charge effect is developed during the process of reflection of the cloud by the earth's field. This electric field is uniform in nature and accelerates all the electrons to approximately the same velocity and energy $\sim 10^4$ ev. The acceleration process does not increase the thermal motions, so that the medium again becomes unstable, and space-charge clouds form and scatter the solar gas into the earth's magnetic field.

Consider a hypothetical magnetic field of uniform strength and direction $\mathbf{H}(0, H_y, 0)$ in the region $z < z_0$ and zero elsewhere. An ion of mass M , velocity $(0, 0, -V)$ would be reflected from the field after penetrating a distance $r = VMc/H_y e$, the charge e being esu. An electron would penetrate a much smaller distance, so that a neutral cloud of density N would tend to

set up an electric field

$$E_z > 4\pi N M V c H_y^{-1} \text{ esu}$$

In the geophysical problem the magnetic pressure may be equated to the rate of destruction of gas momentum, so that

$$H_y^2 = 8\pi N M V^2$$

and then

$$E_z > c H_y / 2 V$$

But such a field will destroy the energy of a heavy ion in a distance

$$r' = M V^2 / 2 e E_z < r (V^2 / c^2) \ll r$$

This result suggests that the heavy ions and electrons are not allowed to separate appreciably under the influence of the magnetic field. The heavy ions must penetrate to nearly the same distance, the heavy ions losing most of their energy to the electric field. Since the electrons penetrate to nearly the same distance (and electric potential is nearly the same) their energy at maximum penetration must be nearly that of the heavy ions at entry, namely $\frac{1}{2} M V^2$. This means that, during the process of reflection, the electrons in a cloud originally moving at 10^8 cm sec⁻¹ should attain speeds $\sim 4 \times 10^9$ cm sec⁻¹.

The problem of the reflection of a gas cloud by a magnetic field has been considered by *Chapman and Ferraro* [1940] and *Ferraro* [1949]. Use will be made here of the latter results, which include the above-mentioned space-charge effect whose potential is found to be

$$\varphi = \frac{e H_y^2 d^2}{2 m c^2} \exp \left\{ \frac{-2(z - z_0)}{d} \right\}$$

where

$$d^2 = m c^2 / 4 N e^2$$

The electric potential falls off exponentially from the innermost surface of the cloud ($z = z_0$) with a scale height of $d/2$. The magnetic field in the cloud⁸ decreases as $\exp(z/d)$. Combining equations 2-4, the maximum value of φ is given by

⁸ The magnetic field in the region $z > z_0$ is zero, as assumed in the simple hypothetical case, but the main argument remains unchanged.

⁷ Since this was written a further paper by *Kahn* [1959] has appeared in which he concludes that the effect of random motions is so devastating that instability is unlikely in any astrophysical situation. The conclusion reached below is between these two extremes, of instability in all cases and no instability at all.

$$\varphi_{\max} = MV^2/e \quad (5)$$

energy gained by an electron corresponds to this value (an induction field is also present) MV^2 , the kinetic energy of the heavy ions in the original cloud. For protons at 10^8 cm sec⁻¹ the maximum electron speed is 4.3×10^9 cm sec⁻¹, their 'temperature' is $\sim 4 \times 10^7$ K. If singly ionized helium atoms are present to control the electromagnetic field, φ_{\max} might have a value as high as 4×10^4 volts or more. If electromagnetic oscillations were present, since ω could increase the effective value of V .

Figure 1 the paths of a single electron and proton in the reflecting region are sketched. The magnetic field is upward from the paper, and the electric field is parallel to Oz . The electron attains its maximum speed, along $-Ox$, near the plane $z = z_0$. Here the protons move along $+Ox$ with a much smaller speed of 2.3×10^8 cm sec⁻¹. Also displayed are plots of the magnetic field strength H and the electric potential φ . When the density of the incident stream is 10^8 cm⁻³, the scale distance d has the value 1.6×10^4 cm.

SCATTERING OF IONS INTO THE EARTH'S FIELD
In the region of thickness d just outside the z_0 plane, two ion streams interpenetrate

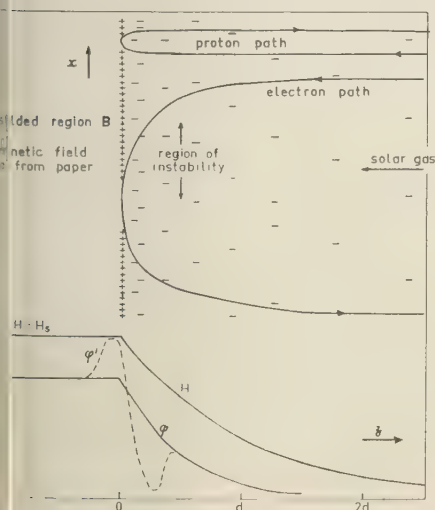


Fig. 1. A schematic diagram of a region of reflection of ions by a compressed magnetic field. The paths of a proton and electron are shown, also plots of the magnetic field H and electric potential φ as given by Ferraro.

with a streaming velocity $\sim 4 \times 10^8$ cm sec⁻¹ corresponding to a thermal energy of the electrons of $\sim 4 \times 10^7$ K. This temperature almost certainly greatly exceeds that of the original solar cloud. It is concluded that, in the region of reflection, conditions are suitable for the growth of space-charge clouds or waves.⁹

Neglecting thermal motions, the dispersion equation, in axes fixed in the proton gas, becomes (see Appendix)

$$\omega_e^2/(\omega - Uk)^2 + \omega_i^2/\omega^2 = 1 \quad (6)$$

The interesting solutions occur where $k \sim \omega_e/U$ [Buneman, 1958] corresponding to clouds of dimension $L \sim \pi U/\omega_e$. Comparing the expression for ω_e (Appendix) with equation 4, we find

$$L \sim \pi d \cdot (U/C)$$

so that in general $L < d$ and the streaming region is thicker than the cloud dimension.

Buneman's calculations show that the maximum rate of growth occurs when the real and imaginary parts of ω (ω_r and ω_i) have values $0.40\omega_e^{2/3}$, $\omega_e^{1/3}$ and $0.69\omega_e^{2/3}$, $\omega_e^{1/3}$. This wave, of form $\exp \omega_i t \exp i(\omega_r t - kx)$ travels one wavelength while growing by a factor 5×10^4 . Most of the region of growth is for slightly longer waves with ω_i/ω_r even larger. Strong growth ($\omega_i = 0.48\omega_e^{2/3}$, $\omega_e^{1/3}$) still occurs when $\omega_r = 0.05\omega_e^{2/3}$, $\omega_e^{1/3}$, corresponding to a wave which grows by a factor $\sim 10^{26}$ while moving one wavelength. The phenomenon might better be regarded as nearly stationary clouds of space charge which grow exponentially with time. They have ample time to develop before they drift out of the region of growth. The growth time is such that $\omega_i t \sim 10$, say, or

$$\begin{aligned} t &\sim 20\omega_i^{-2/3}\omega_e^{1/3} \\ &= 250\omega_e^{-1} \end{aligned}$$

⁹ A second condition for instability is a limitation of the random velocities of the protons. The proton stream of density N may be replaced by an electron stream of density Nm/M without changing the dispersion equation. In this case Bohm and Gross [1949] have shown that the instability condition is that the ratio of thermal to streaming velocities is substantially less than $(m/M)^{1/2}$. This means that in the original cloud the organized velocity must greatly exceed the proton thermal velocity, a condition similar to that for the electrons.

when $N = 10^3 \text{ cm}^{-3}$, $\omega_e = 1.8 \times 10^6$ and $t = 1.4 \times 10^{-4}$ second.

As was mentioned above, the theory of instability is applicable only to small perturbations, but it might be anticipated that a substantial fraction of the streaming energy could be converted into electrical energy. The electric potential ϕ in and near the streaming region would then become irregular as indicated by the dashed line in Figure 1.

It is likely then that in the vicinity of the plane $z = z_0$ space-charge clouds are formed with differences of potential $\lesssim 10^4$ volts. These will grow in periods of order 10^{-4} second, and their size will be of order 10^4 cm for the streaming conditions assumed (corresponding to a major storm). It is inevitable that the formation of such clouds will give rise to hydromagnetic or radio and space-charge waves in the medium between the clouds and the earth. One might anticipate efficient conversion of space-charge cloud energy into wave energy in bands where the appropriate $\omega - k$ relation or dispersion equation of the waves concerned showed frequencies near, but below, 10^4 cycles per second for wavelengths above $\sim 10^4$ cm. The space-charge waves might be important in promoting diffusion and acceleration of ions in the upper atmosphere. The radio waves might reach the earth and account for some of the observed noise in the band below $\sim 10^4$ cps.

For present purposes, a more important result of the space-charge clouds is the scattering of ions into the region $z < z_0$, where some are trapped. One would expect a balance to be maintained between the formation and destruction of the clouds, so that, on an average, the electrons in the region near $z = z_0$ would have the greater energy. Electrons scattered toward the earth might have energies a few times 10^4 ev, and protons somewhat lower. This result is discussed further below.

Let us suppose that a proportion α of all incident ions penetrate behind $z = z_0$ and are trapped. Their mean energy is of order $\frac{1}{2}MV^2$, and in the field H_e they are confined to a layer of thickness ~ 2 mean gyro radii, or, for protons and electrons,

$$D \sim MVc/2H_e e \quad (7)$$

After time t the pressure of this gas is

$$P \sim (2e/3c) \cdot \alpha N V^2 H_e t$$

which equals the magnetic pressure $H_e^2/8\pi \text{ NMV}^2$ after a time $t \sim 2.4 \times 10^7/\alpha V$ (the heavy ions being protons). Suppose that 10 per cent of the incident ions penetrate 'magnetic sheet'; then, if $V = 10^8$ cm/sec, $N = 10^3 \text{ cm}^{-3}$, we have $t = 0.17$ second.

At this stage the thin 'magnetic skin,' which isolates the remainder of the geomagnetic field from the solar gas cloud, becomes inflated and is pushed away from the remainder of the field.

7. CAUSE OF THE MAIN PHASE

If it can be assumed that ions from the solar cloud penetrate into and become connected with the geomagnetic field, it is possible to account for the main phase of a geomagnetic storm. The requirement is that a sufficient number of ions be accelerated rapidly enough *toward* the earth. This is just the reverse of the requirement for the SC, when the ions must be accelerated *away from* the earth. It should also be noted that the acceleration process must continue more or less steadily for the duration of the main phase, that is for a day or so.

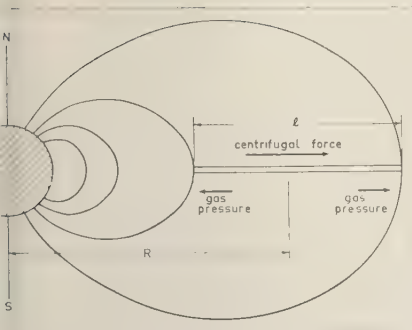
There appear to be only two possible mechanisms, the first involving ions which remain close to the earth for the period concerned. These ions are continually accelerated toward the earth without ever reaching it, by a centripetal force exerted by the geomagnetic field. The ions exert a corresponding centrifugal force.

Centrifugal force theories. In Figure 2 is shown a cylinder representing a 1 cm^2 section through the gas that is exerting the centrifugal force. The axis of the cylinder is radial and lies in the geomagnetic equatorial plane. The cylinder has a length l , and its center is at a distance R from the center of the earth; let N_e , M , and s be the average number density, mass, and size of the ions responsible for the centrifugal force.

Let us first consider the case where all the ions move in circles around the earth in planes parallel to the equatorial plane; there is no spiraling motion. The outward centrifugal force on the column is then

$$F \sim N_e M V_e^2 l R^{-1}$$

This may be compared with the inward pressure NMV^2 of an ion stream of density N and speed V impinging on the outer layers of the geo-



2. Illustrating a column of ionized gas of length l , trapped in the geomagnetic field and producing an outward centrifugal force together with inward and outward pressure.

field during the SC discussed above. When $N, V_e = V$, for reasonable geometries the forces are comparable and so their magnetic effects will be comparable in magnitude, as noted. This is the Chapman-Ferraro ring, which could explain the main phase if the ordered motion of *all* ions in the ring region could be neglected. However, the difficulty of starting the ring motion appears very great [Pidington,

but we now consider the case where the motion of ions in the cylinder is random. The kinetic energy will be divided equally into the three translational, and the centrifugal force will be

$$F \sim \frac{1}{3} N_e M V_e^2 l R^{-1}$$

to the north-south movement along the magnetic field lines. The motion in the other two axes will comprise spiraling around the field lines and other drift motions. Even in the absence of all collisions the ions will, by the action of their diamagnetic effect, exert outward pressure at the ends of the cylinder of length

$$F_z \sim N k T = \frac{1}{3} N M V^2$$

as the penetration into the earth's field is probably large, the ratio $l/R < 1$, and so $F_z > F$: there is a net inward force (due to gas pressure) rather than an outward force as required to account for the main phase.

A third possibility remains, due to the earth's rotation [Parker, 1958b]. However, a comparison of this period of rotation of 24 hours with that

required by the Chapman-Ferraro-Martyn theory of 3 or 4 minutes suggests that the former is not likely to be important in accounting for storms, although it may produce some significant effects.

The geomagnetic 'tail.' It appears to be difficult or impossible to account for the main phase in terms of ions that remain in the vicinity of the earth for a day or so, that is, ions that experience continued centripetal acceleration. The alternative possibility concerns ions that tend to pass the geomagnetic field and move out past the earth's orbit. The momentum available is clearly more than adequate to account for the main phase, and if some lines of force of the geomagnetic field were trapped or frozen into this stream a geomagnetic 'tail' would be formed as shown in Figure 3 [Pidington, 1959] (a similar suggestion has been made by Parker [1958b]).

Ions in the tail are continually accelerated toward the earth by electromagnetic forces in a manner that has been described in connection with gas leaving the vicinity of sunspots [Pidington, 1958]. This must result in a general decrease in the horizontal component of the geomagnetic field and so may account for the main phase.

The difficulty here is to account for the magnetic lines of force becoming connected to the passing cloud. An explanation is attempted in terms of the ion-reflection mechanism shown schematically in Figure 1. It was suggested in section 6 that in the transition region between the solar cloud and the atmosphere a state of balance should exist between two electromagnetic effects. The first effect is the steady-state reflecting field $H(z)$, $\phi(z)$ determined by Ferraro [1952]. The second is a random space-charge electric field $\phi'(x, y, z, t)$ which must inevitably give rise to a set of electromagnetic waves of different types and also cause a scattering of ions, some into the geomagnetic field.¹⁰ If at certain times and places the ϕ' field dominates

¹⁰ The space-charge clouds (ϕ') are nearly stationary in the heavy-ion gas and move rapidly through the main (electron) plasma into which they radiate strongly (owing to their bunching) by the Cerenkov process. Hydromagnetic waves, plasma waves, and extraordinary radio waves are generated because each of these have phase velocities less than the space-charge clouds.

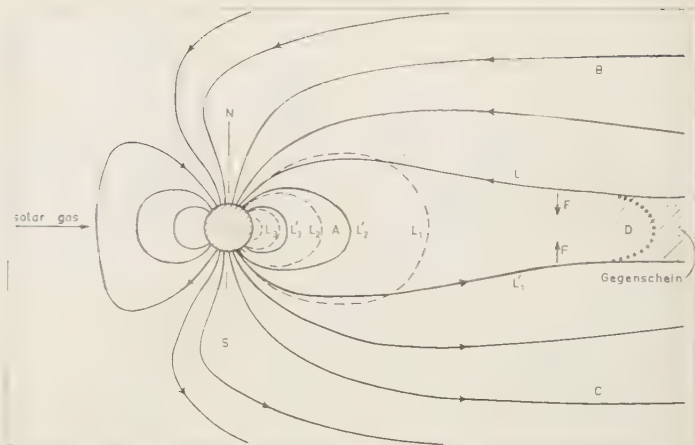


Fig. 3. A schematic diagram of the geomagnetic field during the transition from the SC to the main phase of a storm. Lines of force from all longitudes, but only high latitudes are drawn out to form the magnetic tail pointed away from the sun. The field remaining on the sunlit side is compressed by solar gas impinging on it.

the situation, it must cause a violent frictional effect between the cloud and the field and a 'scattering' of the field by strong hydromagnetic waves. Where the solar ions impinge on the geomagnetic field obliquely, that is in regions where the local time is near 0900 or 1500 hours, the passing gas would sweep away the scattered field.

In the example in section 6 the ϕ and the ϕ' effects each take a period of $\sim 10^{-4}$ second to develop; the same is true for any ion density, so that either effect might be expected to take control at a particular time and place. The formation of the ϕ' field would be favored if a space-charge wave of appropriate length propagated into the region concerned from another region of disturbance. The instability of a double-ion stream increases as the streams are slowed down,¹¹ so that most of the energy of streaming should be convertible to electric field energy. When the ions are slowed their gyroradius is reduced and they become frozen into the field. Subsequently when the ion clouds explode or are blown apart by the electric fields each bunch of ions must carry its own magnetic field. Some of these must be swept away by the passing solar gas. Although the solar gas may tend to

pass through the strand of magnetic field forming a polarization field which cancels induction field $\mathbf{V} \times \mathbf{H}$, the polarization lines of force are carried away [Piddington 1959].

The main phase develops over a period of 12 hours or more, so that because of the earth's rotation magnetic lines of force from many latitudes should contribute to the tail.

Apart from the ϕ' instability, it is possible that purely hydromagnetic instability might be present at the interface of the geomagnetic field and the solar gas [Dungey, 1955; Park, 1958b] and might contribute to the formation of a tail.

8. DISCUSSION OF THE GEOMAGNETIC TAIL

Figure 3 is a schematic representation of the geomagnetic field during the transition from SC to the main phase of a storm. The lines of force in the tail extend to great distances and are connected in the receding gas cloud, which exerts a force sufficient to balance the magnetic tension.

In region A the outermost lines of force of the undisturbed field (L_1) have been pulled out to contribute the lines L_1' to the tail. This reduction in magnetic pressure is transmitted down to earth as a decrease in the horizontal component

¹¹ The minimum size of the growing ion clouds is inversely proportional to the velocity.

all longitudes [Piddington, 1959]; this is indicated here by the lifting of the lines L_2 , L_3 to L_2' , L_3' . There are no sustained mechanical forces available in this region, so that the field must satisfy the equation

$$\text{curl } \mathbf{H} = 0$$

with different boundary conditions from those of the dipole field alone.

The magnetic tail comprises two bundles of lines of force B and C which are oppositely directed. The figure is misleading in showing all lines in the meridian plane; in fact, lines from different longitudes will be twisted around the dipole axis NS and the bundles B and C will be twisted. The flux into the tail during a major storm will be a few times 10^{16} gauss cm^2 [Piddington, 1959], and, if the diameter of each bundle is assumed 10 earth radii, the field strength is about 10^{-3} gauss. Region D is of particular interest because here the gas is compressed by the opposing electromagnetic forces \mathbf{E} . As gas is squeezed out of the region, opposing lines of force meet and connect as shown by the dotted line. The connected lines then contract to re-form the original field. The theory is that of the 'neutral line' given by Sweet [1958] and others. It is suggested below that the gas compressed into region D is responsible for the Gegenschein.

The main mechanical force associated with the tail is exerted by the receding solar cloud. The force density must be equal and opposite to the magnetic tension $H^2/8\pi$ or 4×10^{-8} dyne cm^{-2} for a major storm. This is much smaller than the force needed to stop the solar stream and so may readily be supplied by that stream.

A result of the twisting of lines in the bundles B and C is that a line in B which cuts the earth at one longitude might join a line in C with a different longitude. The result would be a semi-permanent deformation of the outer geomagnetic field. A further deformation should result from the rotation of the earth and of the field within a few earth radii. At greater distances the tail will be controlled by the solar wind, and so a twisting or 'rolling up' of the field must result at intermediate distances.

The process of scattering of ions into the geomagnetic field and the subsequent formation of a geomagnetic tail may explain some effects

other than the main phase of magnetic storms. Those incidental effects include the Gegenschein, auroral and Van Allen radiation, and the diurnal variation of cosmic radiation.

The Gegenschein. The Gegenschein or counter-glow is an elliptical patch of light near the countersolar point, with dimensions about 10° (along the ecliptic) by 6° . It sometimes has a brighter central region and is, on an average, displaced 3° westward of the midnight meridian. Parallax measurements suggest a distance of ~ 20 earth radii and hence dimensions perpendicular to the line of sight roughly those of the earth. Its changing shape before and after midnight suggests that the light originates throughout a column or cone with axis nearly radial from the earth. The emission spectrum is similar to that of the night sky, originating mainly in O and N atoms, and the intensity correlates with auroral activity.

Although the Gegenschein was first observed some 150 years ago a satisfactory explanation has not yet been offered. According to Fesenkov [1950] the radiation comes from a tail of gas driven away from the upper atmosphere, being oval and of width a little greater than 1 earth diameter. Although this promising hypothesis allows several features of the Gegenschein to be explained, it appears inconsistent with more recent upper-atmosphere data. The gas above a level of a few thousand kilometers is almost fully ionized and so is trapped in the earth's magnetic field. It might diffuse out through this field slowly and eventually be swept away by solar radiation, but the tail so formed should be ~ 10 or more earth radii across, which is far too great to be consistent with observations.

The extraordinarily difficult feature to explain is that this part of the night sky light appears to originate in a region of dimensions similar to those of the earth, yet at a distance of ~ 20 earth radii. An explanation of this feature of the Gegenschein follows immediately if it is assumed that the earth usually has a magnetic tail, either due to mild magnetic activity at the time or a remnant of a past magnetic storm. Temperatures in our outer atmosphere are now thought to be as high as 10^6°K , and there is certainly a proportion of ions of much higher energy. Those conditions must ensure a supply of O and N ions throughout the region of the dipole field and hence also of the magnetic tail.

As seen above, the annihilation of magnetic lines of force in the region D (Fig. 3) will result in an accumulation of gas there in the form of a flattened cone or cylinder pointing away from the earth. The light from this region should increase with geomagnetic and auroral activity, as observed.

The main features of the Gegenschein are thus explained in terms of the magnetic tail, and in turn the latter hypothesis is given support. As an extension of these ideas, magnetic tails of other planets, notably Venus and Mars, might be observed in line emission from their atmospheric components.

Auroral and Van Allen radiation. The scattering mechanism described in section 6 should provide ions of energy up to a few times 10^4 ev. It would be expected that electrons would be most plentiful and some might be further accelerated by the various waves scattered from the growing space-charge clouds. During a great magnetic storm enough lines of force should be drawn into the tail to allow the ions to penetrate to latitudes below the auroral zones. At the same time solar gas could stream more or less directly into the field at high latitudes, as shown in Figure 3. Thus some features of auroras might be explained in terms of the magnetic tail, although there is little doubt that other features require localized electric fields.

It is known that on occasions the sun emits ions with energy greater than that of the Van Allen radiation. However, the difficulty of explaining Van Allen radiation in terms of solar ions is the mode of entry into the geomagnetic field to form two sharply defined zones. Simple diffusion would not be likely to form even a single such zone. This difficulty is overcome by the magnetic tail model; the tail will become saturated with electrons that have been accelerated near the earth and with fast solar ions if these are present. When the field resumes its normal form these ions are trapped in a zone at a few ($\gtrsim 3$) earth radii. The outermost ions may then escape by diffusion, leaving a well defined zone.

It is even possible that the inner zone might be accounted for in terms of more energetic ions of a solar origin which penetrated the nontail part of the field while it was exposed during a great storm.

Cosmic-ray variations. The geomagnetic tail

for a major storm might have a cross section of ~ 10 earth radii and a field strength of $\sim 10^4$ gauss. The gyroradius of a cosmic ray moving perpendicular to the field and having an energy of 2 bev is about 10 earth radii, and so the tail would cause a major scattering effect. This is an extreme case, but much weaker tails might be expected to cause noticeable diurnal effects.

As the earth rotates eastward, the lines of force leading into region A should provide a westward twist on the outermost parts of the remaining dipole field. A westward shift of 40° to 45° of the 'cosmic-ray dipole' field has actually been observed [Simpson and others, 1956] and might be explained by the torque exerted by the main-phase field.

APPENDIX

Consider two interpenetrating streams of neutral ionized gas having electron and heavy ion densities $N_{1,2}$, masses m , M , and charge $(esu) \pm e$. The dispersion equation for waves of instabilities of the form $\exp i(\omega t - kx)$ for the four-stream, one-dimensional medium is [Hael, 1949, equations 13a]:

$$\frac{\omega_{e1}^2 + \omega_{i1}^2}{(\omega - U_1 k)^2} + \frac{\omega_{e2}^2 + \omega_{i2}^2}{(\omega - U_2 k)^2} = 1 \quad (8)$$

where

$$\omega_{e1,e2}^2 = \frac{4\pi N_{1,2} e^2}{m}, \quad \omega_{i1,i2}^2 = \frac{4\pi N_{1,2} e^2}{M}$$

and U_1 , U_2 are the stream velocities. Since $\omega_{e1,e2} \gg \omega_{i1,i2}$, the latter terms may be neglected and only the electron streams considered, the positive ions merely constituting a neutralizing background. Equation 8 reveals an instability in the two (electron) stream medium. The kinetic energy of streaming is converted into electric space-charge energy in the form of 'plasma oscillations' or 'growing ion clouds,' depending on the system of axes chosen.

In the analytically simple case of $\omega_{e1} = \omega_{e2} = \omega_e$, $U_1 = -U_2 = U$, any initial space-charge irregularity, however weak, will grow exponentially, provided only the size of the irregularity exceeds $\lambda \sim \pi U / \omega_e$. No oscillation are present; the clouds remain stationary and grow exponentially. When $\omega_{e1} \neq \omega_{e2}$, clouds of different size drift relatively to one another but remain in the velocity range $U_1 \rightarrow U_2$. The

analysis that reveals this cloud growth applies to small perturbations only, but the physical process involved [Piddington, 1958] is so elementary that it seems reasonable to assume that it will continue until the electron streaming is stopped. The original streaming energy will then exist as electric energy of scale λ . For $n = N_2 = 10^3 \text{ cm}^{-3}$ and $U = 10^8 \text{ cm sec}^{-1}$ this scale is about 2 meters.

This situation is brought about in a period of π times ω_e^{-1} , during which the heavy-ion motion is not significantly perturbed. Furthermore, the scale of the electron clouds is much too small ($Uk \sim \omega_e$) to interact with the heavy-ion streams. Thus the electron gas is, for a short period at least, left to its own devices. Its dispersion equation, $\omega^2 = 2\omega_e^2$, does not permit the continued existence of the space-charge clouds or waves which must disintegrate into random motions of single electrons or small groups of electrons. Since energy must be conserved these will have mean velocity $\delta \sim U$. Thus electron interaction has simply converted the streaming velocity into random velocity.

The dispersion equation of the whole gas, with axes fixed in the electron gas, now becomes

$$\frac{\omega_1^2 + \omega_e^2}{\omega^2 - \delta^2 k^2} + \frac{\omega_{i1}^2}{(\omega - U_1' k)^2} + \frac{\omega_{i2}^2}{(\omega + U_2' k)^2} = 1 \quad (9)$$

the term δ being introduced in the manner of Bohm and Gross [1949]. When $\delta = 0$ this shows sets of growing clouds traveling at very nearly the velocities U_1' and $-U_2'$. The interaction between one ion stream and the electron gas is not affected significantly by the second ion stream, and the interaction between the ion streams themselves is too slow to be important. Thus the significant effects present are given by a single heavy-ion stream, the dispersion equation (in axes fixed in the heavy ion gas):

$$\frac{\omega_e^2}{(\omega - Uk)^2 - \delta^2 k^2} + \frac{\omega_i^2}{\omega^2} = 1 \quad (10)$$

When $\delta = 0$, instability occurs in the range $Uk \sim \omega_e \gg \omega$ [Buneman, 1958] as discussed below. However, when $\delta \rightarrow U$ the instability is removed [Bohm and Gross, 1949], allowing the free passage of the ion stream or streams.

It is concluded that neutral clouds may interpenetrate freely once the electron thermal velocity has been raised to equal the streaming velocity of the heavy ions. This heating process requires only a very small proportion of the total kinetic energy of the clouds. The fact that $\delta \rightarrow U$ does not mean that a small cloud emitted from the sun will disintegrate; it will be held together by a positive space charge developed by the loss of a few electrons.

A second situation of interest is that of one electron stream and one heavy-ion stream, constituting an electric current which may be maintained by an applied potential difference or by self-inductance effects. In the latter case the electrons would be heated until $\delta \sim U$. In the former case one might expect the electrons to be accelerated until $U > \delta$, after which their streaming energy would be converted to thermal energy. They would then be accelerated to a yet higher streaming velocity, and so on, to give a continually increasing current. This result would appear to differ from that envisaged by Buneman [1958].

REFERENCES

- Bohm, D., and E. P. Gross, *Phys. Rev.*, **75**, 1864, 1949.
 Buneman, O., *Phys. Rev. Letters*, **1**, 8, 1958.
 Chapman, S., and V. C. A. Ferraro, *Terrestrial Magnetism and Atmospheric Elec.*, **36**, 171, 1931.
 Chapman, S., and V. C. A. Ferraro, *Terrestrial Magnetism and Atmospheric Elec.*, **38**, 79, 1933.
 Chapman, S., and V. C. A. Ferraro, *Terrestrial Magnetism and Atmospheric Elec.*, **45**, 245, 1940.
 Cowling, T. G., *Monthly Notices Roy. Astron. Soc.*, **116**, 114, 1956.
 Dungey, J. W., *Proc. Ionospheric Conf.*, *Phys. Soc. London*, p. 229, 1955.
 Ferraro, V. C. A., *J. Geophys. Research*, **57**, 15, 1952.
 Fesenkov, V. G., *Astron. Zhur.*, **27**, 89, 1950.
 Haefl, A. V., *Phys. Rev.*, **75**, 1546, 1949.
 Hines, C. O., and E. N. Parker, *J. Geophys. Research*, **63**, 691, 1958.
 Hines, C. O., and L. R. O. Storey, *J. Geophys. Research*, **63**, 671, 1958.
 Kahn, F. D., *J. Fluid Mech.*, **2**, 601, 1957.
 Kahn, F. D., *Astrophys. J.*, **129**, 468, 1959.
 Martyn, D. F., *Nature*, **167**, 92, 1951.
 Parker, E. N., *J. Geophys. Research*, **63**, 683, 1958a.
 Parker, E. N., *Phys. Fluids*, **1**, 171, 1958b.
 Parker, E. N., *Phys. Rev.*, **112**, 1429, 1958c.
 Piddington, J. H., *Monthly Notices Roy. Astron. Soc.*, **115**, 671, 1955.
 Piddington, J. H., *Monthly Notices Roy. Astron. Soc.*, **116**, 314, 1956.

- Piddington, J. H., *Australian J. Phys.*, *10*, 515, 1957.
- Piddington, J. H., *Phil. Mag.*, *3*, 1241, 1958.
- Piddington, J. H., The transmission of geomagnetic disturbances through the atmosphere and interplanetary space, *Geophys. J.*, *2*, 173, 1959.
- Shklovskii, I. S., *Astron. Zhur.*, *35*, 557, 1958.
- Simpson, J. A., K. B. Fenton, J. Katzman, D. C. Rose, *Phys. Rev.*, *102*, 1648, 1956.
- Sweet, P. A., *Electromagnetic Phenomena in Chemical Physics*, edited by B. Lennert, Cambridge University Press, p. 123, 1958.

(Manuscript received August 28, 1959;
revised November 2, 1959.)

Occurrence Frequency of Geomagnetic Micropulsations, P_c

J. A. JACOBS AND K. SINNO¹

*Geophysics Laboratory
University of British Columbia
Vancouver, B. C.*

Abstract. An analysis of the occurrence frequency of geomagnetic micropulsations P_c has been carried out using data obtained during the IGY from a world-wide network of stations. From the characteristics of the diurnal occurrence frequency and their latitudinal and longitudinal dependence, the following conclusions are drawn: (1) The occurrence frequency of P_c 's increases as the auroral zones are approached from lower latitudes. Also the hour of the diurnal maximum occurrence appears earlier at high-latitude stations. (2) The occurrence frequency of P_c 's depends not only on local time but also in part on universal time. The universal-time factor affects the modulation of the diurnal occurrence by about 50 per cent. The time of maximum occurrence of P_c 's is about 21 hours GMT in the northern hemisphere. In the southern hemisphere the universal-time factor has opposite phase to that in the northern hemisphere. When the universal-time factor is a maximum in the northern (or southern) hemisphere, the north (or south) geomagnetic pole is about 16 or 17 hours LMT. The GMT dependence derived in this investigation shows about 7 hours' difference compared with Troitskaya's conclusion, which was based on data from several stations in the USSR.

1. INTRODUCTION

In recent years the study of the earth's outer atmosphere has received an added stimulus by the discovery of the existence of remarkably dense ionized particles, and by 'whistler atmospherics.' The use of artificial satellites has also increased enormously the possibility of making direct observations. *Dungey* [1954], *Kato and Watanabe* [1956, 1957a and b], and *Obayashi and Jacobs* [1958a and b] have investigated theoretically the possibility that geomagnetic micropulsations are caused by hydromagnetic oscillations of the earth's outer atmosphere. However, not enough is known about the morphology of these micropulsations to make more precise theoretical interpretations. Several investigations have been made of micropulsations as observed at a single station [*Terada*, 1917; *Holmberg*, 1953; *Angenheister*, 1954; *Duffus and Shand*, 1958], but almost nothing is known about their world-wide morphology. However, *Troitskaya* [1953a and b; 1955] has investigated the GMT dependence of P_c 's and P_t 's using data from several Soviet stations. Such a GMT dependence has not been further investigated, in spite of the fact that her conclusions have been criticized

by many investigators. For these reasons we have analyzed the occurrence frequency of P_c 's using data obtained during the International Geophysical Year from a world-wide network of stations. This analysis has led to a number of conclusions.

2. SOURCES OF THE DATA

For the present investigation, data from rapid-run magnetograms for the period from October 1957 to September 1958 from ten American geomagnetic observatories were used. These stations are listed in Table 1, together with three antarctic stations from which data for October 1957 were obtained. All the data (three components) were photographically recorded by normal-sensitivity rapid-run magnetograms with a recording speed of 4 mm/min.

Additional data from rapid-run magnetograms for the same period (October 1957 to September 1958) were taken from the publications of the International Union of Geodesy and Geophysics. The location of these stations is given in Table 2.

3. OCCURRENCE FREQUENCY OF P_c 'S AT DIFFERENT LATITUDES AND LONGITUDES

By counting the number of hours during

¹ On leave from Radio Research Laboratories, Kokubunji, Japan.

TABLE 1

Station	Abbreviation	Geographic		Geomagnetic	
		Lat.	Long.	Lat.	Long.
Point Barrow	PB	71°18'N	156°46'W	68°.6N	241°.0E
College	Co	64°51'N	147°50'W	64°.7N	256°.5E
Big Delta	BD	64°09'N	145°51'W	64°.4N	259°.0E
Healy	Hy	63°51'N	149°00'W	63°.6N	256°.5E
Sitka	Si	57°03'N	135°20'W	60°.0N	275°.4E
Fredericksburg	Fr	38°12'N	77°22'W	49°.6N	349°.9E
Tucson	Tu	32°14'N	110°57'W	40°.4N	312°.1E
Honolulu	Ho	21°18'N	158°08'W	21°.0N	266°.4E
Guam	Gu	13°27'N	144°45'E	3°.9N	212°.8E
Koror	Kr	07°16'N	134°32'E	3°.3S	203°.5E
Byrd Station		79°59'S	120°01'W	70°.6S	336°.0E
Little America		78°18'S	162°10'W	74°.0S	312°.0E
Wilkes		66°15'S	110°31'E	77°.8S	179°.0E

which Pc 's were observed during the 12 months October 1957–September 1958, Figure 1 was obtained, showing the occurrence frequency of Pc 's at different places in the world. The stations are grouped into four classes according to their geographic location. For clarity the vertical scale for the American stations has been doubled. For comparison, Figure 1 also shows the occurrence frequency of Pc 's at Alma Ata² as determined by Troitskaya from earth current data.

It is rather meaningless to compare the occurrence frequency between different stations, because the sensitivity of the recording equipment varies and because the data are read by

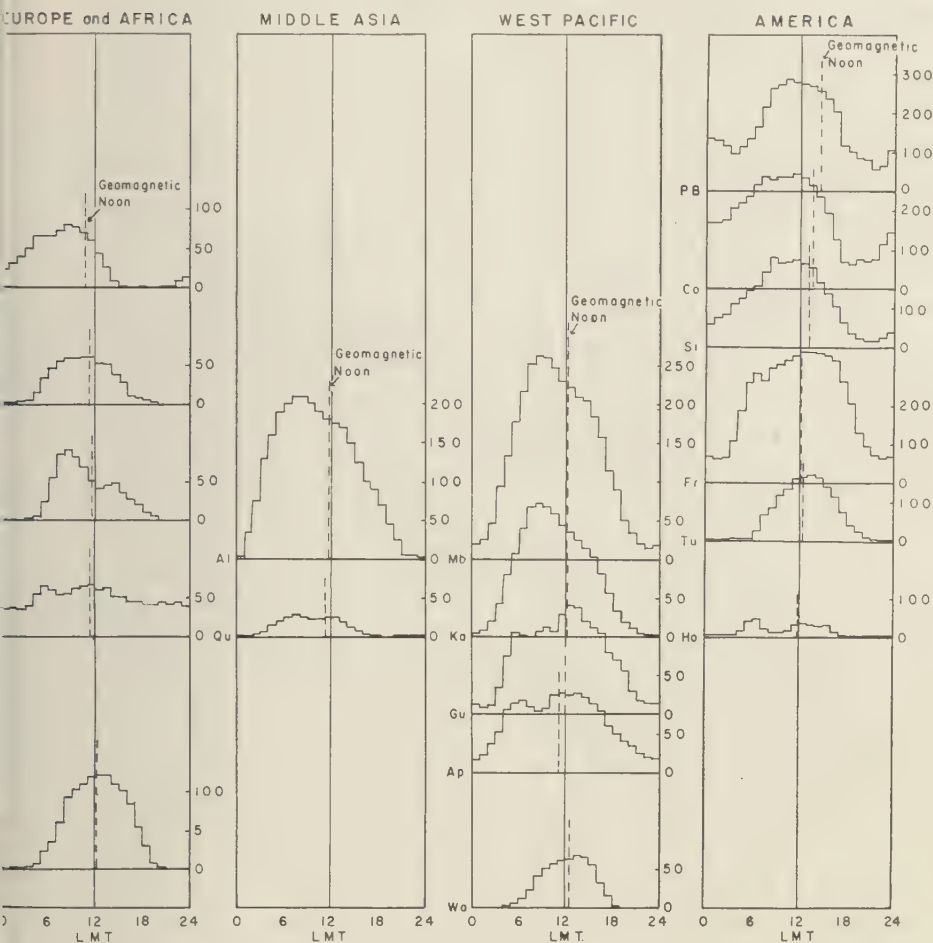
different persons, although it is possible to analyze the form of the diurnal variation. With the American geomagnetic data, however, it is possible to compare the results from the different stations, since the sensitivities of the records at each station are known and since the occurrence of all Pc 's was read by one of the authors (K. Sinno). Microfilm copies of the original records of the rapid-run magnetograms (E trace) were used.

Latitudinal characteristics of Pc 's. The number of hours during which continuous wave trains were recorded with periods in the range from about 10 seconds to 1 minute were counted, regardless of the amplitudes. The percentages of occurrence of Pc 's during the 12 months at each of the American observatories are shown in Figure 2. The points marked '+'

² (Geomagnetic latitude 33°.7N, longitude 150°.5E).

TABLE 2

Station	Abbreviation	Geographic		Geomagnetic	
		Lat.	Long.	Lat.	Long.
Lerwick	Le	60°08'N	01°11'E	62°.5N	88°.6E
Hartland	Ha	51°00'N	04°29'W	54°.6N	79°.0E
Toledo	Tl	39°53'N	04°03'W	43°.9N	74°.7E
Memambetsu	Mb	43°55'N	144°12'E	34°.1N	208°.3E
Kakioka	Ka	36°14'N	140°11'E	26°.0N	206°.0E
Tamanrasset	Ta	22°48'N	05°31'E	25°.4N	79°.6E
Quetta	Qu	30°11'N	66°57'E	21°.6N	139°.7E
Apia	Ap	13°48'S	171°46'W	16°.0S	260°.2E
Hermanus	Hr	34°24'S	19°13'E	33°.3S	80°.3E
Watheroo	Wa	30°19'S	115°53'E	41°.7S	185°.8E

Fig. 1. Diurnal occurrence frequencies of *Pc*'s.

present data uncorrected for sensitivity. After allowing for the different sensitivities (referring everything to 0.2 mm/γ as standard) the points marked 'o' are obtained. The curve through these points shows a pronounced increase as higher latitudes are approached. When this analysis was first carried out, data were available for only 9 months (October 1957 to June 1958). This more limited amount of data indicated a maximum occurrence in the auroral zone. The occurrence frequency decreased slightly from the auroral zone to the pole. The significance of the slight second maximum at the equator is

very doubtful. It was more pronounced in the initial analysis of only 9 months.

The correction for the different sensitivities was made from Figure 3, which was based on the accumulative distribution function derived from data obtained during February 1958 at Fredericksburg (the station with the most sensitive recording equipment). The accumulative distribution is well approximated by the curve $Na = \text{constant}$, where N is the number of hours of *Pc*'s with amplitudes greater than a . On the other hand, the minimum observable amplitude, a_{\min} , is controlled by the sensitivity s ; i.e.

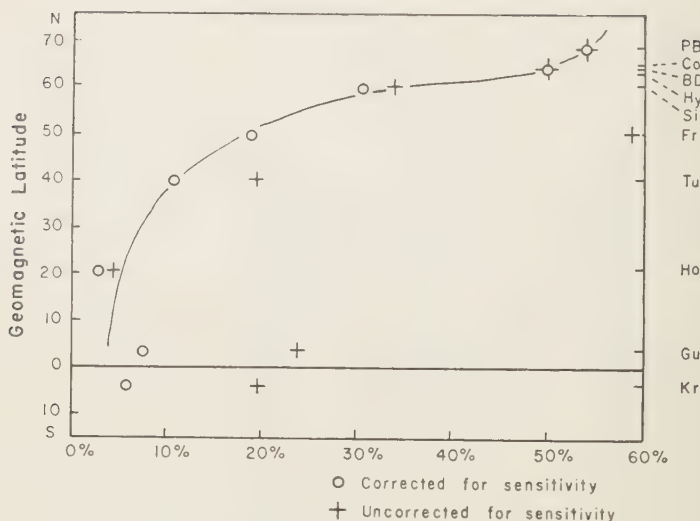


Fig. 2. Percentage occurrences of *Pc*'s.

$$a_{\min} \propto 1/s$$

Thus the total occurrence, N_{tot} , is proportional to the sensitivity s , and it is easy to correct the occurrence frequencies for different sensitivities.

From Figure 1 two characteristics are evident. First, the diurnal peak in the occurrence frequency is earliest in the auroral zone, gradually tending toward the afternoon at lower latitudes. This characteristic is particularly apparent in the records from Europe and North America. Second, in the Western Pacific region, there are two maxima, one prenoon and one postnoon, which predominate in the northern and southern hemispheres, respectively. The limited amount of data from the antarctic observatories confirms this last characteristic. At the antarctic station, Wilkes, south of the West Pacific region, the maximum occurrence of *Pc*'s is at about 15 hours local geomagnetic time in December 1957.

Longitudinal characteristics of Pc's. If the occurrence frequency of *Pc*'s in different regions in about the same geomagnetic latitude is compared, it is found that the maximum occurrence appears earlier at middle and East Asian stations than at American and European stations. For example, the maximum occurrence at Memambetsu in Japan is about 5 hours before that

at Tucson in North America (geomagnetic local time). This suggests that the occurrence frequency of *Pc*'s depends only on universal time as has been stated by Troitskaya, who bases her results on the data from several Soviet stations. In the following analysis, it is more convenient to use 'geomagnetic' local time and 'geomagnetic' universal time. The differences between geomagnetic and normal geographic time at the different observatories are small.

There are not enough observatories to separate statistically the part of the variation that depends on LMT, as has been done for geomagnetic disturbances [Chapman and Bartels, 1951; Nagata and Ono, 1952] and ionospheric disturbances [Martyn, 1953; Sinno, 1953, 1954; Obayashi, 1952, 1954]. However, three observatories, Toledo (74°E) Memambetsu (208°3E), and Tucson (312°E) in approximately the same latitude are located at about equal distances in geomagnetic longitude apart. The separation in longitude of the former two stations is rather large, viz. 133°.6. But the fact that the shape of the diurnal *Pc* occurrence frequency at Alma Ata (150°E) is not so different lends support to the theory that the occurrence frequency of *Pc*'s at places 120° apart may not be too different.

Using data from these three observatories

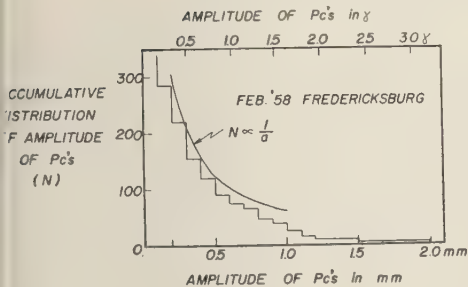


Fig. 3. Accumulative distribution function of amplitude of Pc 's at Fredericksburg, derived from data obtained during February 1958.

equioccurrence frequency patterns of Pc 's (measured in percentages) are shown in Figure 4, where geomagnetic longitude is plotted against geomagnetic universal time. A certain difficulty arises since the data from Toledo and Memambetsu were taken from published numerical figures; i.e. the actual records were not available. Moreover, the records were read by different individuals. However, as explained below, the ratios of the α values from Toledo, Memambetsu, and Tucson were taken as 1.4:0.6:1.0.

Suppose that the occurrence frequency of Pc 's at three equidistant observatories around a circle of latitude can be expressed in the form

$$\begin{aligned} \alpha' A(\tau - \lambda') B(\tau) \\ \alpha'' A(\tau - \lambda'') B(\tau) \\ \alpha''' A(\tau - \lambda''') B(\tau) \end{aligned} \quad (1)$$

where α' , α'' , α''' are factors depending partly on the sensitivity at the three observatories, the functions $A(\tau - \lambda')$, $A(\tau - \lambda'')$, $A(\tau - \lambda''')$ being geomagnetic universal-time factors and $B(\tau)$ a geomagnetic local-time factor. τ is local geomagnetic time and λ geomagnetic longitude, so that $(\tau - \lambda)$ represents 'geomagnetic universal time.' If the function A be expressed as a Fourier series and any differences in α be allowed for, the average of the expressions (1) may be written $\alpha A_0 B(\tau)$, neglecting third- and higher-order terms. This average is purely a local-time factor typical of the 40°N latitude zone of these three stations, and is plotted in Figure 5. It shows an almost symmetrical distribution around noon (local time).

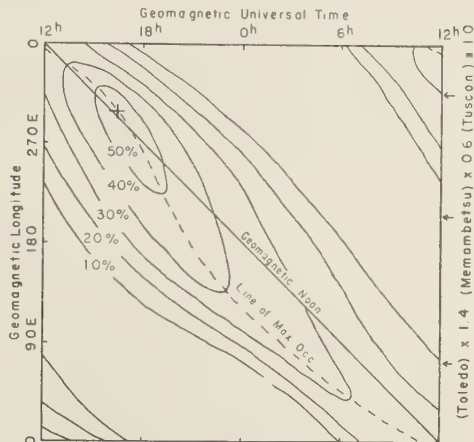


Fig. 4. Equioccurrence frequency pattern of Pc 's.

The geomagnetic universal-time factor $A(\tau - \lambda)/A_0$ can then be obtained for each station by dividing the expressions (1) by the local-time factor $\alpha A_0 B(\tau)$. Its variation is shown in Figure 6. Only daylight time values around noon at the three observatories are shown, since the values at night show more erratic behavior owing to the comparative lack of occurrences. Originally when this curve was drawn with uncorrected values of α it appeared as three discontinuous sections, one for each station. These could be made into a smooth connected curve by using α factors for Toledo, Memambetsu, and Tucson in the ratio 1.4:0.6:1.0. Such ratios also appear reasonable from an inspection of the relative frequency of occurrence at the three stations. The maximum value of the geomagnetic universal-time factor occurs at about 16.5 hours geomagnetic universal time (about 2100 GMT). The universal time factor affects the modulation of the diurnal occurrence by about 50 per cent. The modulation is large, the maximum value being about three times the minimum. The place where Pc 's are most frequent, taking into account both universal- and local-time factors, is located on the west coast of America, and is marked by a cross in Figure 4. The small bump in the curve of Figure 6 at about 3.5 hours geomagnetic universal time (about 0800 GMT) may be the effect of another universal-time factor in the southern hemisphere which may

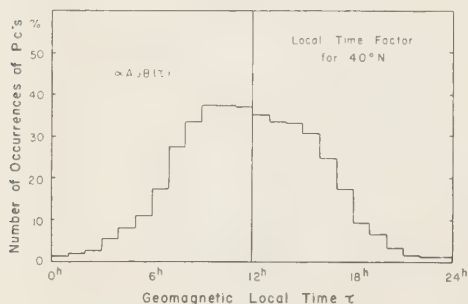


Fig. 5. Geomagnetic local time factor for Pc 's typical of $40^\circ N$.

be antiphased to that in the north, as described at the end of the preceding section.

4. DISCUSSIONS OF THE RESULTS

It is clear that the occurrence frequency of Pc 's depends not only on local time but also on universal time. The universal-time factor may well be connected with the inclination of the geomagnetic to the geographic axis. This investigation supports the suggestion that the location of the exciting origin of Pc 's is on the morning side in the auroral zone. This fact has been anticipated by Kato and Watanabe [1959] on theoretical grounds. Also the morning side in both auroral zones is always excited to some extent, the excitation intensity depending on universal time. The local time of the most excited point appears later in the afternoon as the latitude decreases.

The universal-time factor is a maximum at 21 hours GMT in the northern hemisphere, i.e. when the north geomagnetic pole is about 16 or 17 hours LMT. The universal-time factor in the southern hemisphere may also be a maximum when the southern geomagnetic pole is about 16 to 17 hours LMT. This observation is derived not only from the small bump in the graph of the universal-time factor in the northern hemisphere (see Fig. 6) but also from the occurrence frequency of Pc 's at Hermanus and Watheroo (see Fig. 1). Unfortunately it is not possible to derive the universal-time factor in the southern hemisphere as was done for the northern hemisphere, owing to the lack of data especially from South America. It is worth noting that this GMT dependence differs by

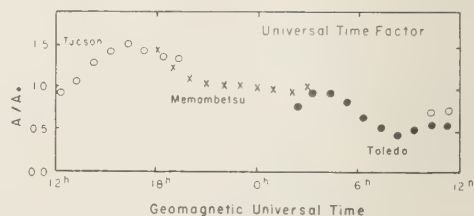


Fig. 6. Geomagnetic universal-time time factor for Pc 's.

about 7 hours from Troitskaya's conclusion.

Micropulsations Pt are now being investigated. It appears that Pt 's have a rather regular diurnal occurrence frequency with respect both to longitude and latitude. Although the investigation is still in its early stages, results to date confirm the observations of Ohchi [1957] that Pt 's show only a LMT and not a GMT dependence.

Data used in this study have been obtained from the geomagnetic stations of the United States and the IGY World Data Center A, Coast and Geodetic Survey, and the International Union of Geodesy and Geophysics. The authors wish to acknowledge the above sources of data. The research was aided by a grant from the U. S. Air Force, Cambridge Research Center, Contract AF #19 (604)-4068.

REFERENCES

- Angenheister, Von G., Registrierung erdmagnetischer Pulsationen, Göttingen 1952/1953, *Gerlands Beitr. Geophys.*, 64, 108-132, 1954.
- Chapman, S., and J. Bartels, *Geomagnetism*, Clarendon Press, Oxford, 1049 pp., 1940, reprinted 1951.
- Duffus, H. J., and J. A. Shand, Some observations of geomagnetic micropulsations, *Can. J. Phys.*, 36, 508-526, 1958.
- Dungey, J. W., Electrodynamics of the outer atmosphere, *Sci. Rept.* 69, Ionos. Research Lab., State Univ. Penna., 51 pp., 1954.
- Holmberg, E. R. R., Rapid periodic fluctuations of the geomagnetic field, I, *Monthly Notices Roy. Astron. Soc.*, Geophys. Suppl., 6, 467-481, 1953.
- Kato, Y., and T. Watanabe, Further study on the cause of giant pulsations, *Sci. Repts. Tôhoku Univ.*, Fifth Ser., Geophys., 8, 1-10, 1956.
- Kato, Y., and T. Watanabe, Studies on geomagnetic pulsation, *Pc*, *Sci. Repts. Tôhoku Univ.*, Fifth Ser., Geophys., 8, 1-22, 1957a.
- Kato, Y., and T. Watanabe, A survey of observational knowledge of the geomagnetic pulsation,

- Sci. Repts. Tôhoku Univ., Fifth Ser., Geophys.*, 8, 157-185, 1957b.
- Kato, Y., and T. Watanabe, Particles of aurorae and geomagnetic pulsations, 1959, in press.
- Martyn, D. F., The morphology of the ionospheric variations associated with magnetic disturbance, *Proc. Roy. Soc. London, A*, 218, 1-18, 1953.
- Nagata, T., and H. Ono, Development of S_D -field with storm-time, *J. Geomag. Geoele.*, 4, 108-113, 1952.
- Obayashi, T., Some characteristics of ionospheric storms, *Rept. Ionosphere Research Japan*, 6, 79-84, 1952.
- Obayashi, T., On the world-wide disturbance of the ionosphere, *Rept. Ionosphere Research Japan*, 8, 135-142, 1954.
- Obayashi, T., Geomagnetic storms and the earth's outer atmosphere, *Rept. Ionosphere Research Japan*, 12, 301-335, 1958a.
- Obayashi, T., and J. A. Jacobs, Geomagnetic pulsations and the earth's outer atmosphere, *Geophys. J., Roy. Astro. Soc.*, 1, 53-63, 1958b.
- Ohchi, K., The temporal (diurnal) distribution of geomagnetic and earth current pulsations, *Mem. Kakioka Mag. Obs.*, 8, 87-92, 1957. (English translation by E. R. Hope, Defense Research Board, Canada, Translation T61J.)
- Sinno, K., On the variation of the F_2 layer accompanying geomagnetic storms, *Rept. Ionosphere Research Japan*, 7, 7-14, 1953.
- Sinno, K., Studies on the disturbances in F_2 layer associated with geomagnetic disturbances, *Rept. Ionosphere Research Japan*, 8, 127-133, 1954.
- Terada, T., On rapid periodic variations of terrestrial magnetism, *J. Coll. Sci., Imp. Univ. Tokyo*, 27, Art. 9, 1917.
- Troitskaya, V. A., Short period (oscillatory) disturbances in the terrestrial magnetic field, *Doklady Akad. Nauk SSSR*, 91, 2, 241-244, 1953a.
- Troitskaya, V. A., Two oscillatory modes of the terrestrial magnetic field and their diurnal GMT cycle, *Doklady Akad. Nauk SSSR*, 93, 2, 261-264, 1953b. (These last two papers have been translated into English by E. R. Hope, Defence Research Board, Canada, Translation T174R).
- Troitskaya, V. A., *Priroda*, 5, 81, 1955.

(Manuscript received July 13, 1959; revised October 26, 1959.)

Daily Normals of the International Magnetic Character Figure, C_4

RALPH SHAPIRO AND FREDERICK W. WARD, JR.

*Geophysics Research Directorate
Air Force Cambridge Research Center
Bedford, Massachusetts*

Abstract. Daily normals (averages) of C_4 have been computed for each day of the year using the 'final' C_4 figures for the 72-year period 1884–1955. The significance of departures of the daily averages from a smoothed continuum is discussed. The observed departures can be attributed to random sampling fluctuations.

It is well known that C_4 , the international magnetic character figure, as well as related geomagnetic indices, has an appreciable semi-annual period with maxima in spring and fall. Unsmoothed daily averages also exhibit this semiannual period. In addition, they reveal occasional large departures from the smooth semi-annual variation. Some uncertainty exists [Maris, 1931; Bartels, 1932; Chapman and Bartels, 1951] about the significance of these departures. Do they represent some real physical phenomenon, or can they be attributed to statistical sampling fluctuations? The purpose of this note is to help clarify this question.

Figure 1 shows each daily 'normal' value of C_4 , computed as the arithmetic average of the corresponding 72 daily values during the period 1884–1955. (No value is given for February

29.) Circled and doubly circled values indicate departures from a smoothed continuum exceeding the 5 and 1 per cent confidence limits, respectively.

The continuum value for any day (designated by a cross in Fig. 1) is defined as the average of C_4 for the period from 2 to 11 days before and after this day. The values of C_4 for the 3-day period centered on the day in question do not contribute directly to the value of the continuum for that day. Since each daily average of C_4 is to be compared with its continuum value, it is believed that the continuum should be constructed so that its value on any one day is reasonably unaffected by the average C_4 for the same day. The values of C_4 on the day before and after the day in question were eliminated in computing the continuum because of

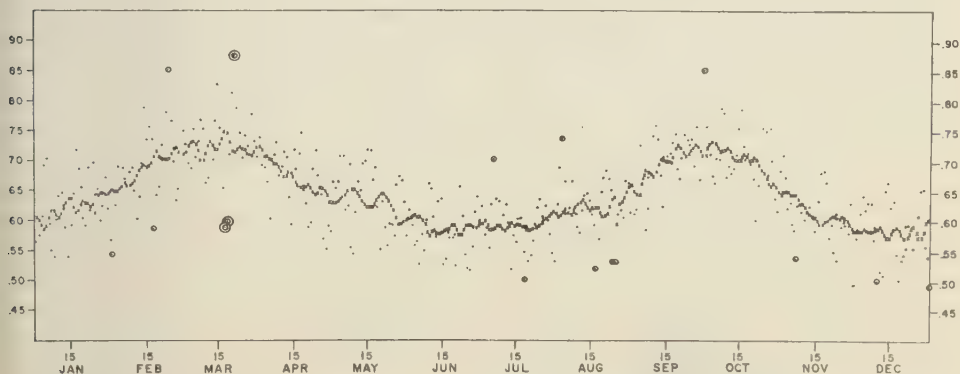


Fig. 1. Daily 'normals' of international magnetic character figure, C_4 (dots), and continuum values (crosses) for the period 1884–1955. Circled and doubly circled values indicate departures from the continuum significant at the 5 per cent and 1 per cent levels, respectively.

the high autocorrelation of C_t at a lag of 1 day.

It can be seen from Figure 1 that the continuum is not completely smooth but exhibits irregularities that tend to be of opposite sign to the departures of C_t from the continuum. These irregularities, which are contributed in part by the elimination of the central 3 days of C_t from the computation of the continuum, tend to increase the magnitude of the departures of C_t from the continuum. As will be shown, however, this bias has no appreciable effect on the nature of the results.

The probabilities that the observed departures (of the daily average values of C_t from their continuum values) are due to chance were determined by the use of Student's t test [Hoel, 1947]; t for any day is defined as the magnitude of this departure divided by the estimated standard deviation of the average value of C_t on this day.

There are 16 days during the year when the departures of the daily 'normals' from their continuum values exceed the 5 per cent confidence limit and 3 days when they exceed the 1 per cent limit. In 365 trials, one would expect the 5 per cent level to be exceeded about 18 times by chance and the 1 per cent level to be exceeded 3 to 4 times. The observed departures from the continuum may therefore be attributed to chance sampling fluctuations in spite of the bias introduced by the definition of the continuum. To remove any doubt about the effect of this bias, a smooth continuum was drawn freehand through the computed continuum values and the significance of the departures of C_t from this smooth continuum was recomputed. In this case, there were 15 days during the year when the departures exceeded the 5 per cent confidence limit and 3 days when they exceeded the 1 per cent limit.

There are other possible measures of the reality of the short-period fluctuations of the daily 'normals' of C_t . A casual examination of Figure 1 shows occasional large changes of the 'normals' within a few days.

Student's t test was again used to determine the likelihood that the observed differences between daily averages were significantly different from chance expectations. The test was applied to all possible differences up to 7 days. In this case, where the test is applied to the differences

TABLE 1. Number of Differences between Daily Average Values of C_t Exceeding Various Confidence Limits

Difference Interval, days	5% Level	1% Level	0.1% Level
1	13	3	0
2	17	1	0
3	26	5	1
4	25	3	1
5	13	2	0
6	11	1	1
7	10	1	0
Total observed	115	16	3
Total expected by chance	128	26	3

between individual daily averages, t is defined as the magnitude of this difference divided by the estimated standard deviation of the average difference between the 2 days in question.¹

Table 1 indicates the results of the t test applied to the differences. Again, the results are close to what would be expected by chance.

A total of 2555 differences was tested. One would expect to obtain by chance 128, 26, and 3 values on the 5, 1, and 0.1 per cent levels respectively. The observed results are very close to these figures.

As an additional check on the representativeness of the results, the data of Figure 1 were divided into halves in two different ways and analyzed separately. The daily average values of C_t computed from the even and the odd years of the period 1884-1955 were compared. Further comparison was made between averages computed from the first and last 36 years of the 72-year period. In each case there was essentially no agreement between the corresponding halves except for the semiannual period. After removal of the semiannual variation by subtracting the continuum values (defined previously), the linear correlation coefficient between the

¹ It should be noted that the standard deviation of a difference

$$(\sigma_{a-b}) = (\sigma_a^2 + \sigma_b^2 - 2\sigma_a\sigma_b\chi_{ab})^{1/2}$$

where χ_{ab} is the linear correlation coefficient between the two quantities.

daily averages for the even and odd years is $+0.079$ and between the first 36 and last 36 years $+0.081$.

The implication of the results of this study is that departures of the long-term daily averages of C_4 from a smooth semiannual variation are due to statistical sampling fluctuations.

Acknowledgment. The authors are indebted to Dr. Walter Orr Roberts, of the High Altitude Observatory, for a suggestion that led to this study.

REFERENCES

- Bartels, J., Terrestrial-magnetic activity and its relation to solar phenomena, *Terrestrial Magnetism and Atmospheric Elec.*, 37, 1-52, 1932.
Chapman, S., and J. Bartels, *Geomagnetism*, Clarendon Press, Oxford, 1951.
Hoel, P. G., *Introduction to Mathematical Statistics*, John Wiley & Sons, New York, 1947.
Maris, H. B., *Annual variations in magnetic storms*, *Phys. Rev.* 37, 1680, 1931.

(Manuscript received October 5, 1959.)

Lunar Daily Variation of Geomagnetic Horizontal Intensity at Alibag

K. S. RAJA RAO

Meteorological Office, Poona, India

Abstract. Following the mathematical development of Chapman and Miller, Tschu has described a practical method of determining the lunar daily variation of geophysical elements. By application of the Chapman-Miller method as described by Tschu, the lunar semidiurnal variation of geomagnetic horizontal intensity at Alibag is worked out for the equinoctial season by making use of the hourly values of the element for the period 1940 to 1944. The expression

$$L(H) = 1.2 \sin(2\tau + 62^\circ) \pm 0.3$$

where $L(H)$ is the amplitude of the lunar variation in the horizontal intensity H and τ is the lunar time reckoned from lower transit of mean moon at Alibag is obtained for lunar semidiurnal variation. $L(H)$ is a maximum when the lunar time is 2 lunar hours past midnight or noon. The solar diurnal variation is also determined up to 4 harmonics. A comparison of these results with similar results for Kodaikanal suggests that solar and lunar variations are independent of each other.

Introduction. The tidal motion in the atmosphere has been the subject of much study—the tides produced by the moon and also those produced by the sun in which thermal as well as gravitational effects enter. In discussing the lunar atmospheric tides, Chapman, to whom a good part of our knowledge of this subject is due, has emphasized the importance of peculiarities in the behavior of such a relatively simple effect as the tidal action. The lunar effect is doubtless simple in contrast with the solar one; for it is purely a gravitational phenomenon, free from the thermal influences that are important in the solar tides. But its magnitude is extremely small in comparison with that of the solar tidal variation, and hence it is very difficult to separate the lunar effect from the dominating solar one. One of the most intriguing peculiarities of the lunar atmospheric tides is the distribution of the phase with latitude. The complexities appearing in the lunar tide are apparently due to the atmosphere itself. For instance, the effect of the moon on geomagnetism is greater during daylight hours than at night. The determination of the effect of the moon at different hours of the day gives valuable information about the phenomenon.

There are two methods of determining the lunar tidal variations in the atmosphere. The

earlier method, called Van der Stok's method, is a fixed-hour method. During a lunation the effect of the moon at all lunar hours is present for each of the solar hours; for example, a particular solar hour t_0 on a certain day the corresponding lunar hour is τ_0 . The next day the moon will move to the east, and the lunar time at the solar hour t will not be τ_0 but less than τ_0 , for the apparent motion of the moon around the earth is slower than that of the sun. The decrease per day will be, on an average, 48.7 lunar minutes or 50.5 solar minutes. The lunar time then is decreasing with increasing dates, for a fixed solar hour t . Therefore, if values for a particular solar hour are taken for decreasing dates, we obtain the hourly values for increasing lunar time. The difference between t_0 and τ_0 determines the lunar phase ν . The same value of ν is repeated after 29.53 days. The phase for each day is given by the value ν , the hour angle of the mean moon at noon. In other words, ν is the number of lunar hours passed after the time of mean noon or it is the lunar time of mean noon. This forms the basis for the lunar harmonic analysis, and it has been extensively used by many workers in determining the lunar tides in the atmosphere.

Chapman and Miller [1940] gave a general statistical method for determining the lunar

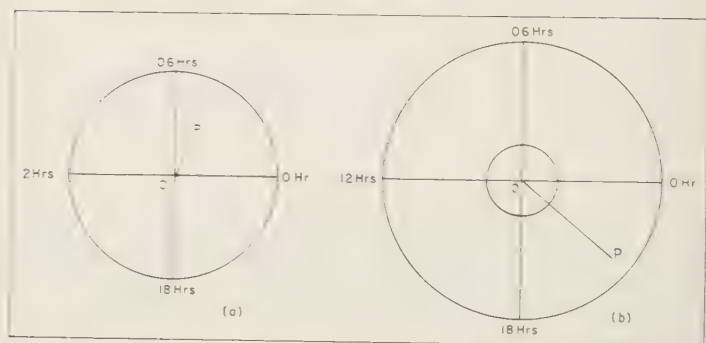


Fig. 1. Harmonic dials representing lunar diurnal variation at (a) Alibag and (b) Kodaikanal.

variations. This is the fixed-age method. The lunar groups of days are chosen according to the hour angle of the mean moon at Greenwich mean noon. This method enables one to calculate not only the lunar components but also the solar and the lunisolar components. Correction for nonperiodic variation can be applied, and the disturbed days can easily be omitted. Calculation of the probable error is also possible. The practical application of the method and the computational procedure have been fully worked out, as explained above, by *Tschu* [1949]. Following *Tschu*, the geomagnetic data for Alibag have been analyzed in the present paper.

Choice of data. The hourly values of geomagnetic horizontal intensity at Alibag for the 5-year period 1940–1944 have been used. The minimum part of the sunspot cycle has been chosen in order to minimize the effects of sunspot activity.

The coordinates of Alibag are geographic latitude, $18^{\circ}38'N$; geographic longitude, $72^{\circ}52'E$; geomagnetic latitude, $9^{\circ}30'N$; geomagnetic longitude, $143^{\circ}06'E$. The hourly values for the equinoctial *E* season, March, April, September, and October, have been used in the analysis. All international disturbed days, as selected by the International Association of Geomagnetism and Aeronomy, have been omitted.

Method of computation. The computations were made as described previously [*Raja Rao and Sivaraman*, 1958]. The expression for the semidiurnal lunar variation, which is the main component, is

$$L(H) = 1.2 \sin(2\tau + 62^{\circ}) \quad (1)$$

Probable error of the above determination has been worked out according to the method outlined by *Tschu* and later by *Chapman* [1952] and it is found to be 0.3.

It is possible to calculate from these computations the solar daily harmonic components for $p = 1, 2, 3, 4$. The solar variation can be expressed as

$$\begin{aligned} S(H) = & 16.4 \sin(t + 68^{\circ}) \\ & + 6.6 \sin(2t + 206^{\circ}) \\ & + 2.4 \sin(3t + 231^{\circ}) \\ & + 1.0 \sin(4t + 377^{\circ}) \end{aligned} \quad (2)$$

From the results of the calculation it is seen that the solar harmonics are large; probable errors have therefore not been calculated.

Discussion. The lunar geomagnetic variation for Kodaikanal (geomagnetic latitude -0.6°) has been determined by *Raja Rao and Sivaraman* [1958], and the expression

$$L(H) = 3.32 \sin(2\tau - 39^{\circ}04') \quad (3)$$

has been obtained for the lunar semidiurnal wave. The lunar variations at the two stations Kodaikanal and Alibag, are represented in the harmonic dials in Figure 1. It is seen that the amplitude of the lunar variation in the magnetic equatorial region is nearly 3 times as large at a higher latitude.

The S_q variation at Kodaikanal is given by the expression

$$\begin{aligned} H) = & 19.2 \sin(t + 48^\circ) \\ & + 14.6 \sin(2t + 29^\circ) \\ & + 7.8 \sin(3t + 80^\circ) \\ & + 2.2 \sin(4t + 283^\circ) \quad (4) \end{aligned}$$

The amplitudes of the lunar and solar variations at Alibag and Kodaikanal represented by expressions (1), (2), (3), and (4) do not bear any relationship. The lunar and solar tidal variation may therefore originate from different causes.

REFERENCES

- Chapman, S., The calculation of the probable error of determinations of lunar daily harmonic component variations in geophysical data: a correction, *Australian J. Sci. Research, Ser. A*, 5, 218-222, 1952.
- Chapman, S., and J. C. P. Miller, The statistical determination of lunar daily variations in geomagnetic and meteorological elements, *Monthly Notices Roy. Astron. Soc., Geophys. Suppl.*, 4, 649-669, 1940.
- Raja Rao, K. S., and K. R. Sivaraman, Lunar geomagnetic tides at Kodaikanal, *J. Geophys. Research*, 63, 727-730, 1958.
- Tschu, K. K., On the practical determination of lunar and luni-solar daily variations in certain geophysical data, *Australian J. Sci. Research, Ser. A*, 2, 1-24, 1949.

(Manuscript received April 15, 1959; revised September 8, 1959.)

Survey of Number of Solar Flares Observed During the International Geophysical Year¹

HELEN W. DODSON AND E. RUTH HEDEMAN

*The McMath-Hulbert Observatory
The University of Michigan
Pontiac, Michigan*

Abstract. According to the McMath-Hulbert Observatory working list of world-wide flares there were more than 6700 solar events during the IGY for which at least one observer estimated the importance to be ≥ 1 . According to reports to World Data Centers the flare patrol was gratifyingly complete and uniform. However, examination of the starting times of the reported flares reveals evidence for a serious lack of homogeneity in the IGY flare data. Approximately twice as many flares, ≥ 1 , were reported per hour from 05^h to 16^h UT as during the remainder of the Universal Day. This strong dependence on UT exists in the flare data for each month of the IGY.

Examination of the patrol hours and flare reports for May and July 1958 for each of the 19 H α cinematographic patrol stations shows that the principal photographic patrol stations fail to confirm the very high rate of flare occurrence for 05^h to 16^h UT that appears in the total IGY data. Visual observations between 05^h and 16^h UT, apparently constitute the principal source of inhomogeneity in the flare data. For the hours 17 to 04 in the Universal Day the IGY flare reports are based primarily on the cinematographic patrol.

Comparison of the number of flares and subflares indicates that in the IGY data for the hours 05 to 16 UT many flare events are listed as flares that probably would have been classified as subflares by the principal cinematographic observers.

INTRODUCTION

For the 18 months of the International Geophysical Year there was world-wide cooperation of solar astronomers to keep the sun under observation at all times in order to provide as complete coverage as possible for solar flares and other transient solar phenomena. A flare patrol was considered as having been in operation if at least one observer was examining the solar disk less frequently than once every 5 minutes. At most of the flare-patrol stations, the sun was photographed in H α radiation once every 30 seconds. Forty-four stations participated in the flare patrol, and flares observed at these stations were reported to each of three IGY Solar World Data Centers.

All reports of flares to the World Data Centers for which at least one observer estimated the importance of the flare to be ≥ 1 have been organized, chronologically, at l'Observatoire de

Meudon, and the resulting lists of reports have been prepared for publication in forthcoming issues of the *Quarterly Bulletin of Solar Activity*. During the IGY, many observers saw or photographed the same flares and sent their numerous, and often differing, measurements and evaluations to the World Data Centers. All this information is given in the flare tables of the *Quarterly Bulletins*.

In order to summarize or study solar-flare activity during the IGY, it is necessary to prepare a working list of flares, in which the tabular entry for each flare is the best evaluation that can be made, flare by flare, of all the separate reports of flares in the *Quarterly Bulletins*. Such a working list of world-wide IGY flares has been prepared by one of us for use at the McMath-Hulbert Observatory on the basis of prepublication pages of the *Quarterly Bulletin* flare tables kindly supplied by Mme d'Azambuja and M. Olivieri, of l'Observatoire de Meudon. For flares observed and reported by only one station, the information in the *Quarterly Bulletin* has been copied verbatim in the McMath-Hulbert Observatory working list of IGY flares. For flares

¹ Assisted in part by the U. S. National Committee for the IGY of the National Academy of Sciences and the National Science Foundation through Project NSF Y/22.3/306.

with more than one report in the *Quarterly Bulletin*, the entry in the McMath-Hulbert Observatory working list is a careful evaluation of all the individual reports.

TOTAL NUMBER OF FLARES OBSERVED DURING IGY

The McMath-Hulbert Observatory working list of IGY flares shows that more than 6700 different solar flares were observed between July 1, 1957, and December 31, 1958. For all these flares at least one observer estimated the importance to be at least as great as 1. Table 1 and Figure 1 show the distribution of the IGY flares by importance and by month.² From the foregoing data it is obvious that during the IGY a solar flare was not a rare event. The number reported varied irregularly from month to month, being greatest in September and October 1957 and least in January, February, and November 1958. In no month were fewer than 200 flares reported.

It should be noted that, for 835 of the flares reported in the *Quarterly Bulletin* and included in our lists, the importance of the flare events, as judged on the basis of all the data and notes in the *Quarterly Bulletin*, was considered to be 1—, or that of a subflare. Note also that the *Quarterly Bulletins* and the tabulations here presented exclude the many thousands of solar events originally and unambiguously reported as subflares (flares of importance 1—).

COMPLETENESS OF IGY FLARE PATROL

The success of the world-wide effort to provide complete flare coverage for the 18 months (13,176 hours) of the IGY can be appraised from the data in Table 2 and Figure 2A. According to reports to World Data Centers, and the summarizing charts in the *Quarterly Bulletins*, a flare patrol was 'in operation' for more than 12,500 hours, or 95 per cent of the entire 18 months of the IGY. There was no month for

² Flares in Tables 1 and 3, and Figures 1 and 3, labeled '(1—*)' were reported as of importance ≥ 1 by at least one observer and are included in the *Quarterly Bulletin*. On the basis of all the data in the *Quarterly Bulletin*, they have been evaluated as of importance 1— and are so indicated in the McMath-Hulbert Observatory working list of IGY flares.

TABLE 1. Number of Solar Flares Observed Reported Each Month during the International Geophysical Year

Month	(1—)*	Importance			Total
		1	2	3 and 3+	
1957 July	28	334	39	5	378
Aug.	36	320	25	2	363
Sept.	50	368	53	10	481
Oct.	71	436	21	2	529
Nov.	35	275	18	2	330
Dec.	37	298	27	0	362
1958 Jan.	13	196	23	1	233
Feb.	9	215	24	0	248
March	38	352	54	4	448
April	35	295	14	1	345
May	53	313	18	2	426
June	91	254	32	1	378
July	90	340	31	3	464
Aug.	86	335	33	4	458
Sept.	71	305	28	1	395
Oct.	27	279	23	0	329
Nov.	29	189	14	2	234
Dec.	36	285	20	1	342
Total	835	5389	497	41	6272

* See footnote 2.

which the coverage dropped below 88 per cent and for 9 months (half of the IGY) the coverage was being 'patrolled' for flares more than 96 per cent of the time. Furthermore, the flare patrol provided almost equally good coverage throughout each hour of the day. (See Table 2 and Figure 2B.) It was only for 01^h and 02^h UT that

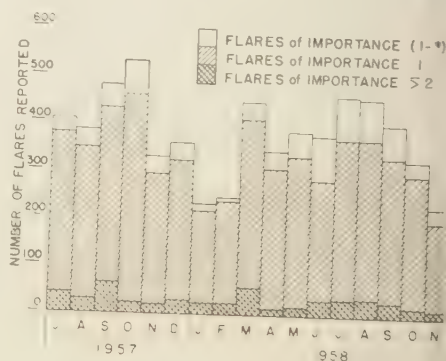


Fig. 1. Distribution of flares by month during IGY
* See footnote 2.

TABLE 2. Summary of Flare Patrol during IGY, Based on Charts in *Quarterly Bulletin* (prepublication data)

Month	Patrol by Months		UT	Patrol, by UT, for 18 Months (549 days)	
	Total Hours in Operation	Per Cent Complete		Total Hours in Operation	Per Cent Complete
7 July	730	98.1	0	501	91.3
Aug.	732	98.3	1	489	89.1
Sept.	710	98.6	2	479	87.2
Oct.	657	88.3	3	496	90.3
Nov.	669	92.3	4	513	93.4
Dec.	661	88.8	5	523	95.3
8 Jan.	695	93.3	6	500	91.1
Feb.	630	93.8	7	530	96.5
Mar.	691	92.9	8	546	99.5
Apr.	695	96.5	9	546	99.5
May	721	96.9	10	546	99.5
June	713	99.0	11	540	98.4
July	723	97.2	12	529	96.4
Aug.	706	94.9	13	538	98.0
Sept.	702	97.5	14	537	97.8
Oct.	716	96.2	15	532	96.9
Nov.	683	94.8	16	532	96.9
Dec.	688	92.4	17	523	95.3
			18	536	97.6
Total	12,522	95.0	19	538	98.0
			20	529	96.4
			21	518	94.4
			22	501	91.3
			23	500	91.1
			Total	12,522	95.0

the patrol was less than 90 per cent complete, and from 07^h to 20^h UT it was in operation more than 95 per cent of the time.

The completeness of the flare patrol necessarily influenced the number of flares observed and reported to the IGY World Data Centers. However, according to the charts in the *Quarterly Bulletin*, the flare patrol was so nearly uniform and complete that variation in flare coverage should have introduced differences of more than 10 or 20 per cent in the number of flares observed.

HOMOGENEITY OF IGY FLARE DATA

All who observe the sun are aware of the many difficulties inherent in the measurement of the area and intensity of flares and in the estimation of their importance. In order to promote homogeneous observations, standards for estimation of flare importance and guides for differ-

entiation between flares and subflares were drawn up by Commission 11 of the International Astronomical Union and distributed to all IGY flare-observing stations. Before IGY flare observations are used for correlation with other solar or geophysical phenomena, it is desirable to try to appraise the degree of homogeneity that actually was achieved in the IGY flare program.

If the flare patrol during the IGY was comparably complete for all hours, and if the basic standards of flare identification and evaluation were similar at the various observing stations, the starting times of the flares should form a nearly uniform distribution with respect to the 24 hours of the 'Universal Day.' However, Table 3 and Figure 3 show that this is not the case. Instead, the number of flares in the IGY data is heavily dependent on Universal Time, and the departure from uniformity is so great that

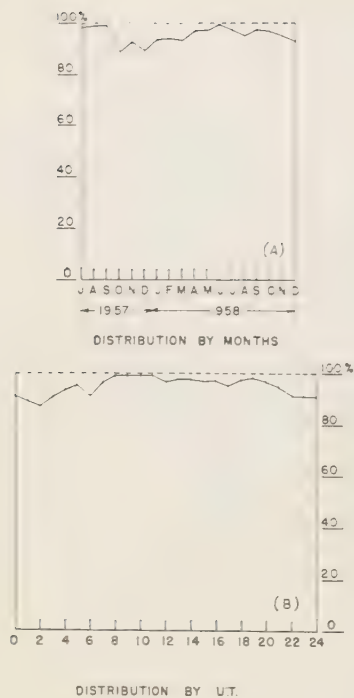


Fig. 2. Percentage of time during which IGY Flare Patrol (based on charts in the *Quarterly Bulletins*) was in operation.

it requires explanation before the IGY flare data can be used satisfactorily.

According to the McMath-Hulbert Observatory working list of IGY flares, approximately twice as many flares, ≥ 1 , were reported per hour from 05^h to 16^h UT as during the remainder of the day. For each hour from 07 to 10 UT, inclusive, the number of reported flares was greater than 450 and corresponded to a rate of flare occurrence greater than 0.8 flare per hour. For the 12 hours of the day starting at 17^h UT, the number of reported flares represented the occurrence of only 0.2 to 0.3 flare per hour. (See Fig. 3.) These large differences in IGY observational data can result from (a) unsuspected differences in the completeness of the flare patrol, (b) systematic differences in the standards of flare evaluation, or (c) a 'periodicity' of $24 \pm$ in solar-flare activity. The last explanation should be ruled out unless it is shown that the differences in flare reports are not

TABLE 3. Number of Flares Reported as Stationing or First Observed for Each Hour (UT) during the IGY

UT	Importance				Total
	(1-)*	1	2	3 and 3+	
0	16	173	16	4	209
1	21	136	13	1	171
2	11	143	18	3	175
3	14	133	22	1	170
4	16	162	14	2	194
5	16	244	25	1	286
6	9	291	27		327
7	41	369	44	2	456
8	39	442	31	2	514
9	42	425	40	5	512
10	33	416	29	2	480
11	37	340	21	2	400
12	53	297	31	2	383
13	93	330	22	2	447
14	90	254	20	3	367
15	97	210	12		319
16	68	185	14	4	271
17	32	134	9	2	177
18	24	124	12	1	161
19	25	122	18	1	166
20	25	91	18		134
21	13	112	16	1	142
22	7	112	12		131
23	13	144	13		170
Total	835	5389	497	41	6722

* See footnote 2.

amenable to explanations on other grounds. The subsequent discussion indicates that it is necessary to call upon a solar periodicity to explain the large differences in the reported number of flares.

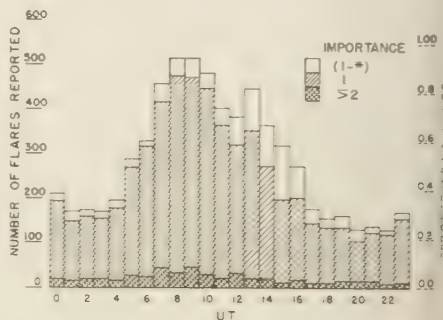


Fig. 3. Distribution of flares in IGY Data according to time of start or first observation.

* See footnote 2.

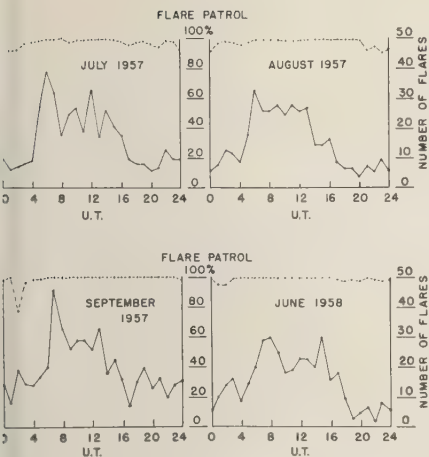


Fig. 4. Number of flares in IGY data, versus time for 4 months with nearly complete flare patrol. Solid line shows number of flares; dashed line shows completeness of flare patrol.

Completeness of flare patrol. Our evaluation of the completeness of the IGY flare patrol given above, based on *Quarterly Bulletin* charts of patrol reports, indicates that inequalities in the patrol do not introduce differences of more than 10 per cent in the number of flares from hour to hour, but the actual difference in the IGY flare data is of the order of 200 per cent or more. The summaries also show that the strong dependence of the number of flares on Universal Time exists in the data for each month of the IGY. That this dependence is not a simple effect of inequalities in the flare patrol is apparent in the data for July, August, and September 1957 and June 1958—months in which the flare patrol was 98 per cent complete (see Fig. 4). For each of these months for which there is apparently no question of the adequacy of the flare patrol, the number of reported flares is apparently a function of Universal Time.

In order to check for the possible effect of unexpected inequalities in the flare patrol, we have compared independent data on the hours of flare coverage provided by the cinematographic $H\alpha$ flare-patrol stations. It should be pointed out that, in theory, if at least one good flare-patrol station is in operation at all times, it is sufficient to provide 'complete' flare coverage.

The hours of cinematographic $H\alpha$ flare patrol have been determined for the months of May

and July 1958, from data supplied by World Data Center A. The numbers of flares reported by the cinematographic stations for these months have been derived from data in the *Quarterly Bulletins*. The information is summarized in Table 4 and presented graphically in Figures 5 and 6.

Nineteen stations were participating in the $H\alpha$ photographic flare patrol in these months. Inter-comparison of the hours of cinematographic flare patrol by all the stations shows that during these 2 months the coverage provided by these stations was as low as 75 per cent only at 02^h and 10^h UT (see Fig. 5A). For the 2 months as a whole, the photographic patrol was in operation more than 88 per cent of the time.³ The cinematographic patrol became most uniform and complete for hours 12 to 20 UT, the very hours during which the number of flares in the IGY data was systematically declining (see Fig.

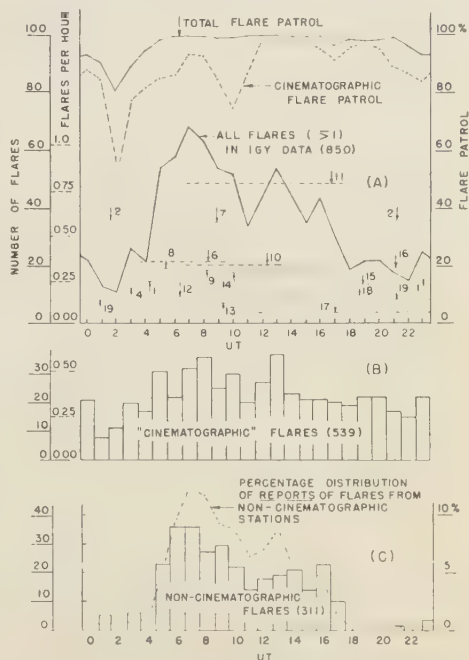


Fig. 5. Study of data for $H\alpha$ cinematographic flare patrol for May and July 1958. See Table 4 for identification of separate stations by code number.

³ In addition, Mount Wilson Observatory conducted a cinematographic calcium flare patrol and reported the recorded flares to the IGY.

TABLE 4. Data for H α Cinematographic Flare-Patrol Stations, May and July 1958

Station* and Identifying Number	Approximate UT of Local Noon	Approximate Number of Hours of Flare Patrol			Number of Flares Reported			Flares per Hour of Patrol		
		May	July	Total	May	July	Total	May	July	Total
1 Sydney	02	76	60	136	18	14	32	0.24	0.23	0.24
2 Voroshilov	03	53	28	81	21	26	47	0.40	0.93	0.66
3 Alma Ata	07	17	11	28	14	15	29	0.82	1.36	1.09
4 Tashkent	07	77	73	150	14	16	30	0.18	0.22	0.20
5 Baku	09	42	17	59	4	2	6	0.10	0.12	0.11
6 Abastumani	09	66	56	122	19	24	43	0.29	0.43	0.36
7 Nizmir†	10	41	66	107	14†	40†	54†	0.34	0.61	0.48
8 Crimea	10	68	80	148	9	37	46	0.13	0.46	0.30
9 Kiev/o	10	68	47	115	22	14	36	0.32	0.30	0.31
10 Capetown	11	76	106	182	21	39	60	0.28	0.37	0.32
11 Uccle‡	12	42	84	126	28‡	62‡	90‡	0.67	0.74	0.71
12 Meudon	12	72	120	192	10	19	29	0.14	0.16	0.15
13 Herstmonceux	12	28	58	86	1	9	10	0.04	0.16	0.10
14 Ottawa	17	42	52	94	7	20	27	0.17	0.38	0.27
15 U.S.-NRL	17	117	154	271	18	41	59	0.15	0.27	0.21
16 McMath	18	208	162	370	48	69	117	0.23	0.43	0.33
17 Climax	19	11	131	142	...	8	8	...	0.06	0.06
18 SacPeak	19	214	173	387	23	39	62	0.11	0.23	0.17
19 Honolulu	22	158	106	264	7	27	34	0.04	0.25	0.16
Summation and average		1476	1584	3060	298§	521§	819§	0.20	0.33	0.27

* In addition, Mount Wilson conducted a photographic calcium flare patrol; May, 214 hours; July, 242 hours.

† Excludes 4 flares in May and 3 in July which were reported for times outside of cinematographic patrol.

‡ Excludes 8 flares in May and 2 in July which were reported for times outside of cinematographic patrol.

§ Note that this is a summation of *reports* of flares, not a count of different flares.

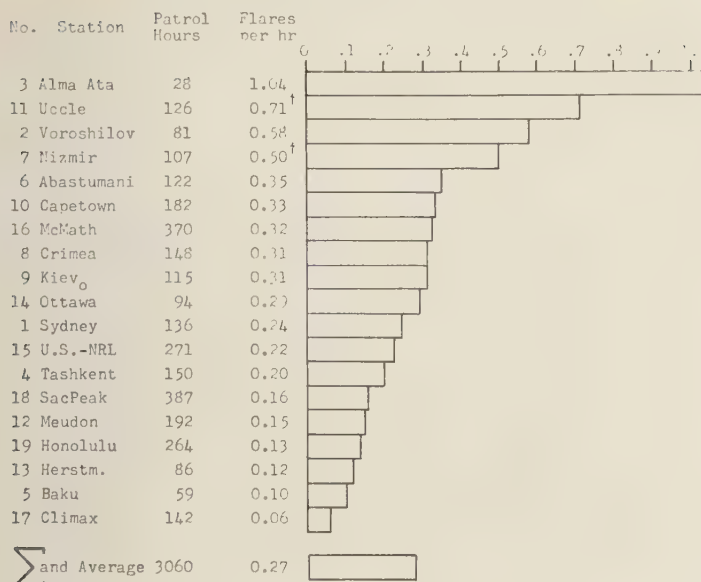
5A). This information provides additional evidence that the strong dependence on Universal Time of the number of flares in IGY lists cannot be explained as a simple result of the existence or absence of the flare patrol for successive hours of the Universal Day.

Differences of standards in flare reporting. It now becomes necessary to try to investigate whether there is any evidence for systematic differences in the solar events that were reported to the IGY World Data Centers as 'flares of importance at least as great as 1.' It is not a question of right or wrong, but merely a matter of homogeneity in the data from many observers in widely separated stations often using very different types of instruments.

All who examine the multiple reports and measurements in the *Quarterly Bulletin* for a single flare become aware that there can be wide differences in measurement, as well as in opinion,

regarding solar flares. If the complete photographic records of the flares with discordant reports are examined, however, it is usually possible to understand why various starting times have been given, why the reported maxima differ so greatly, and even why some observers disagree greatly about the size of the flare in question. Solar flares are often so complex that they do not yield readily to simple codification and measurement.

Furthermore, all who have examined excellent photographic H α records of the sun taken near the time of solar maximum realize that there is hardly an instant when some small, flare-like phenomenon is not in progress on the solar disk. Most of these solar features are smaller than a millionths of the solar hemisphere and therefore, according to the IAU-IGY recommendations, fall into the category of *subflares*, unless they are unusually bright. The task of distinguishing



[†]These rates exclude 10 flares reported by Uccle and 7 flares reported by Nizmir for times outside of the cinematographic patrol

6. Flares (≥ 1) reported per hour of observation at $H\alpha$ cinematographic flare-Patrol Stations, May and July 1958.

between flares and subflares was one of the most difficult assignments given the IGY flare observers, and could be accomplished only on the basis of actual measurement or long experience in solar observation. This problem, or difficulty, existed for visual as well as for photographic observers, since many subflares are easily visible in spectrohelioscopes and in small solar telescopes using $H\alpha$ filters.

In an effort to explore the homogeneity of the reports from different stations, we have determined the 'number of flares reported per hour of observation' for May and July 1958 for each of the 19 cinematographic flare-patrol stations. (See Table 4 and Fig. 6.) For May, the average number of flares per hour of observation at the cinematographic stations was 0.20, for July 0.33, and for the 2 months combined 0.27. Examination of the data for the individual stations shows that, although for certain stations there were large and significant departures from the averages, for *all* the $H\alpha$ cinematographic stations with more than 150 hours of flare patrol during May and July 1958 the rates of flare occurrence were between 0.33 and 0.13 flare per hour of observation. The above information is presented

graphically in Figure 5A, in which we have shown by horizontal dashed lines the average number and/or rate of flare occurrence for May and July 1958, for each of the separate $H\alpha$ cinematographic stations, for the hours of the Universal Day generally patrolled by the respective stations. Data for stations with less than 62 hours of total patrol (less than an average of 1 hr/day) have been excluded from this diagram. It can be seen from this presentation that, in general, the principal $H\alpha$ cinematographic flare-patrol stations fail to confirm the very high rate of flare occurrence between 05^h and 16^h UT indicated by the total IGY flare data for May and July 1958.⁴

⁴At Sac Peak, Honolulu, and Climax flare areas are not corrected for projection effects. Application of the cosine correction to the Sac Peak flare areas for May and July 1958 changed 53 subflares to flares of importance ≥ 1 , gave a total of 115 flares, and raised the number of flares per hour of patrol from 0.16 to 0.30. Thirty-eight of the 53 cases had been reported as flares ≥ 1 by other stations and are in the Quarterly Bulletin. The remaining 15 cases constitute a potential increase of only 8% in the number of flares between 16^h and 23^h.

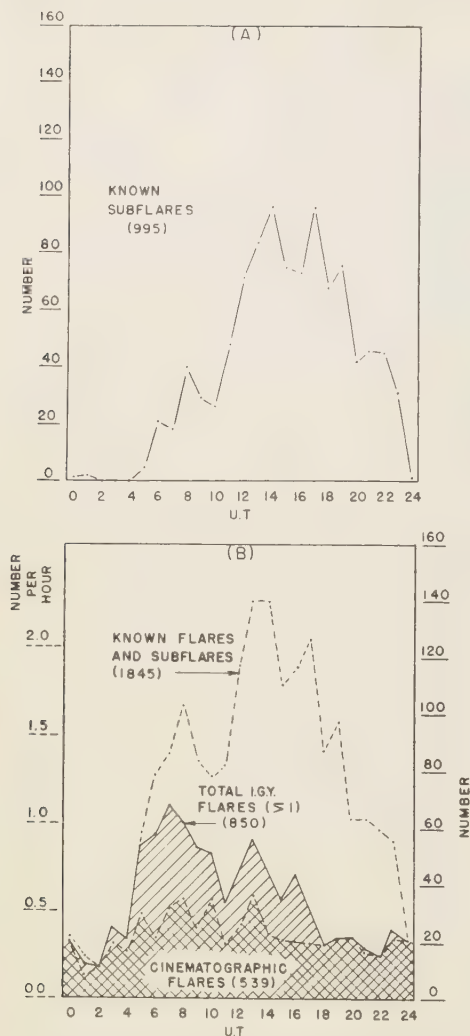


Fig. 7. Summary of known flares and subflares for May and July 1958. Sources of data: *Flares*, McMath-Hulbert Observatory working list of IGY flares based on data in the *Quarterly Bulletins*; *subflares*, summary of additional solar events reported as subflares in *F-Series Bulletins* of CRPL of U. S. National Bureau of Standards.

In addition a count has been made of all solar events reported as flares of importance ≥ 1 by at least one cinematographic patrol station during May and July 1958 (see Fig. 5B). Although the counts are higher than average between 05^h

and 13^h UT, they likewise fail to confirm the very high rate of flare occurrence indicated by these hours by the total IGY data for the months in question.

The above discussion suggests that the great difference in number of flares in the IGY data for the different hours of the Universal Day stems primarily from systematic differences between cinematographic (basically photographic) and 'other' (primarily visual) observations. If this is correct, there should be more reports of flares from noncinematographic stations for the hours with the very large number of flares than for the other hours of the day. Figure 5C shows that there were. Furthermore, it shows that from 18^h to 04^h UT the IGY flare lists for May and July 1958 are based almost entirely on cinematographic records, but between 05^h and 17^h the reported flares represent a combination of visual and photographic data.

Because photographic flare observers could measure and evaluate the various flare events with care and without haste, it is possible that they may have distinguished more closely between flares and subflares than visual observers. Consequently the IGY data for 05^h to 17^h UT may include as flares a large number of events that during the 'photographic' hours of the day would have been classified as subflares.

SUBFLARES

It is unfortunate that complete world-wide data for subflares are not available. On the basis of our best information,⁵ we have determined the number of different solar events that were reported to World Data Centers as subflares, and for which no other station estimated the importance to be ≥ 1 . The distribution is given in Figure 7A. These subflares were reported primarily by the H α cinematographic patrol stations, especially those observing in the latter half of the Universal Day. The counts show that the many solar brightenings smaller than 100 millionths of the hemisphere were well recorded on the patrol films, and were noted and reported to World Data Centers as subflares.

In Figure 7B, we have brought together the counts of flares and subflares for May and Ju-

⁵ CRPL-F series, which reports primarily on flares from 11^h to 24^h UT.

According to the evidence here presented, we conclude that, during these 2 months, there were events on the sun at the rate of 0.3 to 0.4 per hour for which all observers would agree on the classification 'flare of importance 1 or greater,' and an additional and equal number of events for which some observers would report 'flare' and others 'subflare.' In addition, there were thousands of small brightenings for which all observers would agree with the subflare classification.

CONCLUSION

Finally, if the foregoing discussion is valid, it must be concluded that the IGY flare data as reported in the *Quarterly Bulletins* is non-homogeneous with respect to Universal Time. In the hours 05 to 16 UT many solar events were listed as flares that probably would have been classified as subflares by photographic observers. If the lack of homogeneity in the IGY flare lists is realized by those who use the data for ionospheric studies or for studies in other

disciplines, there should be no serious difficulty. It is also clear that it is difficult to distinguish between flares and subflares except on the basis of careful measurement. The difference is primarily one of area, and has been introduced only because of the practical difficulty, or near impossibility, of reporting all small brightenings of a flare-like nature that occur on the sun near the time of maximum activity. Those who study flare-associated phenomena during the IGY should bear in mind the strong dependence on Universal Time that exists in the IGY flare data. Whenever there is an interest in small-scale solar events, flare lists should be supplemented with lists of subflares, or primary $H\alpha$ records of the sun should be examined for the intervals in question.

We acknowledge with pleasure the assistance of Miss Sara Higgins and Mr. Frederic Stewart in the preparation of the data used in this paper.

(Manuscript received October 3, 1959; revised November 2, 1959.)

The Nightly Variation of Auroras at a Subauroral Station¹

JOSEPH W. CHAMBERLAIN AND HELENE M. THORSON

*Yerkes Observatory
University of Chicago
Williams Bay, Wisconsin*

Abstract. The observing records of E. E. Barnard and more recent patrol photographs of the sky have been studied for an indication of a nightly variation in the occurrence of auroras at Yerkes Observatory. There appears to be a maximum probability in the neighborhood of local magnetic midnight, as at stations closer to the auroral zone. Some implications to auroral theory are discussed.

Introduction. The nightly variation of auroral occurrences is, like much of the general auroral morphology, poorly understood. The behavior may be different at different locations, but there are some indications from earlier work that it depends on the auroral form. Vegard [1912] first found some consistency in observations at several stations from the First Polar Expedition, 1882-1883, when he showed a rather general tendency for the maxima to occur around magnetic midnight, although there was a wide scatter in the corresponding geographic latitudes.

There have been suggestions of a less pronounced secondary maximum in the early morning hours at some stations but not at others [Lambert, 1931]. A dependence on geomagnetic latitude for various auroral forms has also been found by Fuller [1933, 1935; also see Fuller and Smith, 1937; Currie and Edwards, 1934; Davies, 1935, 1950; Jacka, 1953; and Elvey, Schindler, Hessler, and Noxon, 1955].

Observations at Yerkes Observatory. Our study has been based mostly on the observing notes kept by E. E. Barnard from 1897 through 1922 at Yerkes Observatory (geographic, 43°34'N, 88°33'W; geomagnetic, 53°N, 24°W). Some of his records and descriptions of auroras have been compiled from his notebooks and published by Barnard [1902, 1910], but for the later years

we have had to depend exclusively on his personal records.

In an earlier statistical study of low-latitude auroras, Meinel, Negaard, and Chamberlain [1954] analyzed Yerkes Observatory records over a 55-year period. Unfortunately, we have been unable to utilize these records after Barnard's death in 1923, as usually no details were given on the time of occurrence. We have supplemented the Barnard records, however, with auroras recorded by an all-sky camera from 1952 through 1956.

The nights have been divided into 1-hour intervals, and an occurrence was counted if at any time during the hour an aurora was observed in any part of the sky. No attempt has been made to correct the statistics for cloudiness, and it has not been possible to divide the occurrences into categories of brightness or auroral form. The totals have been grouped around the equinoxes and solstices. A few occurrences near morning or evening twilight have been discarded in plotting the results (Fig. 1); the night-time intervals shown include only the hours during which an aurora could be detected on any night in each 3-month period. Figure 1 includes 489 auroras giving 1457 hourly occurrences from Barnard's observations and 47 auroras giving 244 occurrences recorded with the camera.

A maximum frequency occurs near the middle of the night during all seasons. The maxima are not especially pronounced, especially near the winter solstice, and no significance can be attached to the different times of maximum

¹The research reported in this paper was supported in part by Geophysics Research Directorate of the Air Force Cambridge Research Center, Research and Development Command, under contract AF 19(604)-3044.

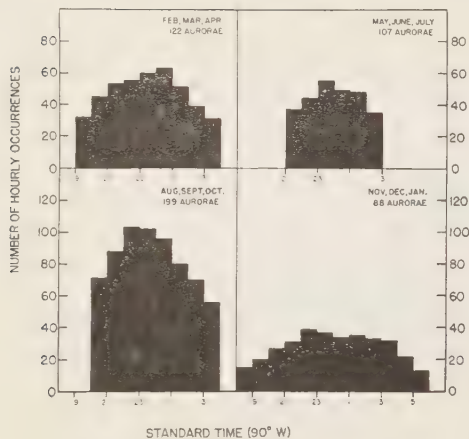


Fig. 1. Distribution of hourly occurrences of auroras. In this figure an aurora is represented in an hourly interval if it was observed at any time during that interval.

occurrence in the various monthly groups.

The results have been plotted against standard time for the 90th geographic meridian west, but this time is very nearly the same as geomagnetic time at this station. The geomagnetic pole (which refers to the centered-dipole approximation) has an azimuth (or magnetic declination) of $6^{\circ}26'$ east of north for Yerkes Observatory. Magnetic midnight (when the sun crosses the lower magnetic meridian) is never later than 26 minutes after geographic midnight (at the summer solstice) and only 17 minutes later at the equinoxes and 9 minutes at the winter solstice. The hour angle of the true sun generally differs from the standard time, owing to the displacement of the station from the standard meridian and to the equation of time. At Yerkes the true sun crosses the lower meridian 22 minutes before midnight in late October and the early part of November; usually the difference is much less than that.

Figure 2 shows the distribution of duration times for these auroras. The figure indicates a maximum probability for auroras of short duration, less than an hour. But it is necessary to bear in mind that this distribution may be profoundly affected by the visual threshold for auroral detection and does not necessarily indicate a probability distribution for the duration of particle bombardment (or other causes). In

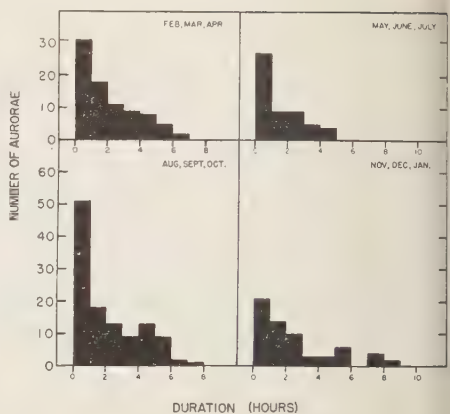


Fig. 2. Distribution of duration times of auroras. If an aurora was observed during 3 (say) hourly intervals, as shown in Figure 1, it was assumed to have a duration of 2 to 3 hours. Auroras that were first observed during evening twilight or that extended into morning twilight were not counted.

preparing this figure we have discarded all auroras that were present at evening twilight or that were cut off by the appearance of dawn, as there is no way of telling how long the auroras would have been detectable if not obscured by sunlight. In so doing we have eliminated most of the displays of long duration and relatively few short ones, thereby overemphasizing the scarcity of long displays.

That the shorter displays are the most frequently observed may be due simply to the facts that their frequency of occurrence decreases rapidly with increasing brightness and that an aurora varies in brightness through the night. Thus a long-lived, very faint aurora might be observed for only a short time when it brightens, whereas fluctuations in a strong display would not matter. The great auroras that become quite bright and extend far from the auroral zone often persist for many hours and even possibly for more than one night. A study of the relation between the brightness and duration of displays would be of interest but is not possible with these visual data.

Discussion. Considerable interest is attached to the problem of nightly variations of auroral and particle bombardment from the standpoint of auroral theory. (For a review of the theories of aurora and of the motions of charged parti-

in magnetic fields see Chamberlain [1958].) The IGY programs of all-sky photography and polar spectroscopy may be especially valuable in this problem by establishing the variations in proton and electron bombardment separately. The appearance of auroras on the night side of the earth has always aggravated the problem of auroral theory, and only Störmer's theory has attempted a precise explanation of this fact. The momenta of auroral primaries are so low, however, that Störmer's explanation—that particle orbits in the outer magnetic field may have such a large radius of gyration that they are deflected around to the dark hemisphere—cannot be valid for auroras in the auroral zone and lower latitudes. Single auroral particles will usually be nearly adiabatically invariant in the magnetic field out to several earth radii, and they would be expected to follow nearly a single line of force from their entry into the field to the upper atmosphere. Meek [1955] and Nikolskii [1957] have found that magnetic disturbances in the arctic show daily variations, the magnetic time of maximum disturbance being a function of magnetic latitude. They have interpreted these results in terms of the precipitation curve for Störmer particles, which, on a long-time statistical basis, may be considered to emanate from a wide range of latitudes. Malville [1959] has shown that the time of maximum occurrence of some antarctic auroras (inside the auroral zone) during the IGY depended in a similar fashion on their geomagnetic latitude. Incidentally, aside from the fundamental difficulties in Störmer's theory, it is not clear how low-latitude results could be fitted into the picture, since his theory requires extraordinary conditions for an aurora to penetrate beyond the auroral zone. There are, however, at least three plausible alternatives, each requiring an *ad hoc* hypothesis, for explaining particle bombardment in the dark hemisphere. (a) The particles from the sun are not emptied into the earth's field exclusively on the day hemisphere, but the solar stream as a whole bends around to the nightward side. (b) Particles in the outer realm of the terrestrial field are affected by an appreciable electric, as well as the magnetic, field, and so drift immediately to the nightward side, where they strike the atmosphere. (c) Particles are

injected on the day side, but, the vast majority being magnetically reflected before entering the atmosphere, they drift and soon become dispersed over all longitudes. In the last picture the outer radiation (Van Allen) belt would presumably constitute potential auroral primaries and not just rejects. Some mechanism is then required to release the trapped particles on the night side.

None of these hypotheses has yet been developed adequately. But it is reasonable to expect that an appropriate auroral theory, when it appears, will offer a natural explanation of the nightly variation of auroras.

REFERENCES

- Barnard, E. E., Observations of the aurora made at the Yerkes Observatory 1897-1902, *Astrophys. J.*, **16**, 135-144, 1902.
- Barnard, E. E., Observations of the aurora made at the Yerkes Observatory 1902-1909, *Astrophys. J.*, **31**, 208-233, 1910.
- Chamberlain, J. W., Theories of the aurora, *Advances in Geophys.* (H. E. Landsberg and J. Van Mieghem, eds.), **4**, 109-215, Academic Press, New York, 1958.
- Currie, B. W., and H. W. Edwards, Summary of some auroral height-measurements and observations at Chesterfield, Canada, *Terrestrial Magnetism and Atmospheric Elec.*, **39**, 293-297, 1934.
- Davies, F. T., The diurnal variation in magnetic and auroral activity at three high latitude stations, *Terrestrial Magnetism and Atmospheric Elec.*, **40**, 173-182, 1935.
- Davies, F. T., Visual auroral observations in Canada, 1943-47, *Bull. Intern. Assoc. Terrest. Magnetism and Elec., Intern. Union Geodesy and Geophys.*, **13**, 255-273, 1950.
- Elvey, C. T., H. Leinbach, H. Hessler, and J. Noxon, Preliminary studies of the distribution of auroras in Alaska, *Trans. Am. Geophys. Union*, **36**, 390-394, 1955.
- Fuller, V. R., Auroral observations at the Alaska Agricultural College and School of Mines for the years 1931-1932, *Terrestrial Magnetism and Atmospheric Elec.*, **38**, 207-238, 1933.
- Fuller, V. R., A report of work on the aurora borealis for the years, 1932-1934, *Terrestrial Magnetism and Atmospheric Elec.*, **40**, 269-275, 1935.
- Fuller, V. R., and E. H. Bramhall, Auroral research at the University of Alaska 1930-1934, *Univ. Alaska Misc. Publs.*, **3**, 1937.
- Hulburt, E. O., On the diurnal variation of the aurora polaris, *Terrestrial Magnetism and Atmospheric Elec.*, **36**, 23-40, 1931.
- Jacka, F., The southern auroral zone as defined by the position of homogeneous arcs, *Australian J. Phys.*, **6**, 219-228, 1953.

- Malville, J. M., Antarctic auroral observations, Ellsworth Station, 1957, *J. Geophys. Research*, 64, 1389-1393, 1959.
- Meek, J. H., The location and shape of the auroral zone, *J. Atmospheric and Terrest. Phys.*, 6, 313-321, 1955.
- Meinel, A. B., B. J. Negaard, and J. W. Chamberlain, A statistical analysis of low-latitude aurorae, *J. Geophys. Research*, 59, 407-413, 1954.
- Nikolskii, A. P., The world-wide distribution of magneto-ionospheric disturbance and aurorae, *Doklady Akad. Nauk SSSR*, 115, 84-87, 1957.
- Vegard, L., On the properties of the rays producing aurorae borealis, *Phil. Mag.*, 23, 211-23, 1912.

(Manuscript received September 28, 1959.)

Southern-Hemisphere Observations of Sodium Emission throughout Twilight

B. J. O'BRIEN¹

*Antarctic Division
Department of External Affairs
Melbourne, Australia*

Abstract. A report is given of preliminary measurements in Melbourne and Antarctica of the twilight and nightglow sodium emissions. These observations were part of the program of the Australian National Antarctic Research Expeditions (ANARE). A photometer with a filter of 5 Ångström bandwidth was used. Continuous measurements of the intensity were begun when the earth's shadow was about 20 km above the earth. The height of the sodium layer causing the twilight emission was about 90 km in both Melbourne and the Antarctic. The ratio of twilight intensity to the nightglow intensity at the end of twilight varied between 15 and 50. It is suggested that further study of the ratio may clarify several problems associated with the nightglow.

Introduction. Observers in the northern hemisphere have found that both twilight and nightglow sodium D emissions vary considerably in their intensity from night to night. The average intensities in winter, however, are 5 to 10 times greater than the average intensities in summer [Munten, 1956; Roach, 1955]. Rather limited observations in the southern hemisphere have indicated that the twilight intensity was relatively constant from April to September [Mayaud and Robley, 1954]. The desirability of further southern-hemisphere observations has been pointed out on many occasions.

Whereas variation of twilight intensity is caused by a variation in the amount of sodium present at appropriate altitudes, variation of nightglow intensity might result in part from atmospheric cause and in part from a variation in the excitation. By studying the ratio (P) of twilight intensity to the nightglow intensity at the end of twilight, the contribution of the excitation mechanism can be isolated. In the past this has not been done.

In this note we describe our measurements of sodium in Melbourne, together with twilight measurements in the Antarctic. These observations are preliminary to more systematic studies to be undertaken by ANARE in Melbourne and at Lawson, Antarctica, throughout 1960.

Experimental details. The sodium twilight emission was measured on the ratemeter of a portable photometer [O'Brien and de la Harpe, in press] on three evenings in February 1959 at latitudes between 60° and 66°S during the southern voyage of the ANARE relief ship, *M. V. Magga Dan*. In June and July 1959, measurements were made in Melbourne with the apparatus modified so as to feed into a pen recorder. Limited nightglow measurements were made in Melbourne in the summer of 1958-1959, and more extensive ones in June and July 1959.

A 2-inch-square multilayer interference filter from Baird-Atomic, Inc., was mounted in front of a lens. The field stop of the lens was a part of the photocathode of an EMI photomultiplier type 6256A with low dark current. The field of view was circular and 1.8° in diameter.

The filter used in this experiment had a peak transmission of 28 per cent at 5893.6 Å for a beam of collimated light at normal incidence ($\theta = 0^\circ$). The bandwidth (measured at half-maximum intensity) was 4.6 Å. At $\theta = 10^\circ$ the peak transmission was at about 5860 Å. At $\theta = 0^\circ$, the transmission of D_2 and D_1 was about 8 and 13 per cent respectively. At $\theta = 10^\circ$, the transmission of each was negligible.

Measurements discussed here are of the intensities only in the zenith. For evening twilight observations, the intensities at $\theta = 0^\circ$ (I_0) and $\theta = 10^\circ$ (I_{10}) and the dark current were measured at least once a minute from just after

¹ Now at Department of Physics, State University of Iowa, Iowa City, Iowa.

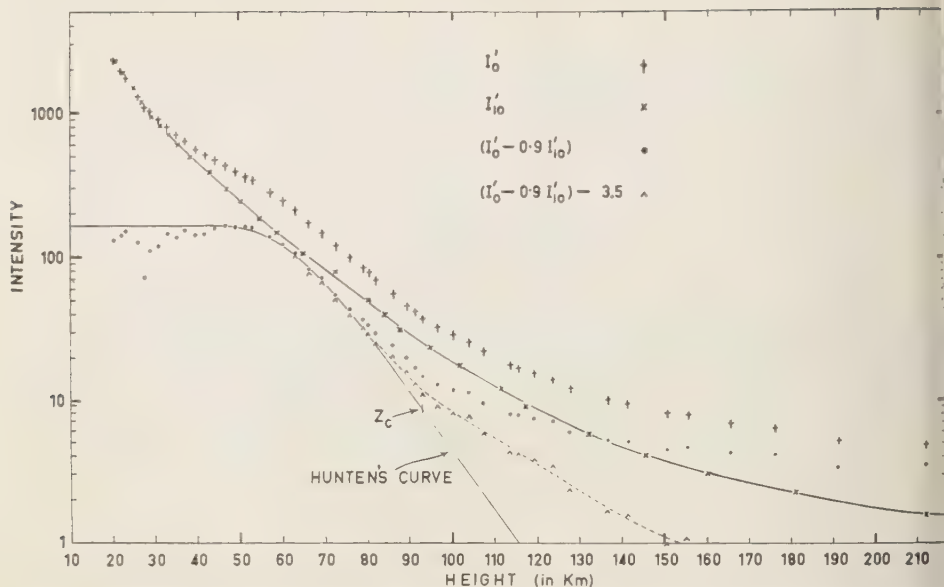


Fig. 1. Intensity of sodium emission and solar continuum as a function of height of earth's shadow. For explanation, see text.

sunset until I_0 and I_{10} diminished to their steady night-time values. Similar observations in reverse order were made for morning twilight.

During the measurements in Melbourne, I_0 was recorded continuously except for rapid measurements of I_{10} and dark current about every minute. A time constant of 1 second and a chart speed of 0.5 mm/sec were used. When observations were continued throughout a night, the filter was rocked by a synchronous motor from $\theta = 0^\circ$ to $\theta = 10^\circ$ and back once each minute. From such sweeps the ratio of the intensities of the D lines (i.e., D_2/D_1) can be found.

The dark current was subtracted from each measurement of I_0 and I_{10} , the correction being less than 10 per cent even for weak nightglow emissions. The corrected intensities I'_0 and I'_{10} were plotted against the height (h) of the unrefracted shadow cast by the center of the sun. A typical good set of measurements is shown in Figure 1. The units of intensity are arbitrary.

Analysis. The filter at $\theta = 0^\circ$ passed the sodium emission and also a contribution from the solar continuum. The solar contribution was reduced to some extent by placing a Polaroid in front of the filter. Nevertheless, the back-

ground or solar contribution to I_0 was considerable, and allowance must be made for it.

Let the intensity of the sodium emission at any time be I_{Na} . Also, suppose that the intensity of the background spectrum passed at $\theta = 0^\circ$ is a constant (x) times the background spectrum passed at $\theta = 10^\circ$, viz. I'_{10} .

Then, for any height h ,

$$I_{Na} = I'_0 - xI'_{10}$$

For small values of h , i.e. just after sunset $I_{Na} \ll I'_0$ and $I_{Na} \ll I'_{10}$, and hence x can be found.

I'_0 and I'_{10} are plotted in Figure 1, and we find that $x = 0.9$. We fit a smooth curve to I'_{10} , and then plot $(I'_0 - 0.9I'_{10})$, where I'_{10} is the value read from this curve for the height at which I'_0 is measured.

Hunten [1954] has derived theoretical curves with which we may compare our plot of $(I'_0 - 0.9I'_{10})$ versus h . He considers the sodium to have a maximum concentration at a height Z_c , and an exponential distribution with a scale height of 7.5 km above and below Z_c . By taking account of (i) single scattering only, (ii) refraction, (iii) the finite size of the sun, and (iv) absorption by ozone, Hunten was able to derive

theoretical curve (curve *B*), which fitted satisfactorily his measurements of sodium intensity versus h for $20 < h < 80$. Curve *B* is plotted by Hunten by considering all heights as being measured from the height Z_c as zero. In order to find Z_c for a given set of measurements, one then fits curve *B* to the measurements, and Z_c is the value of h that coincides with the zero height of curve *B*.

Accordingly, we fitted Hunten's expression for his curve *B* to the points of $(I_0' - 0.9I_{10}')$ and found it to be a good fit for $30 < h < 80$ km. In this case Z_c is 92 km.

The points for $h < 30$ km scatter about 50 per cent from Hunten's curve, but are somewhat closer to the curve of *Blamont and others* [1958]. Assuming Hunten's curve to be correct, since I_0' and I_{10}' are at least 10 times I_{Na} for $h < 30$, it is apparent that the actual measurements of I_0 and I_{10} are accurate to at least 50 per cent. This accuracy should then hold for all measurements with $h > 20$ km, provided that the intensity of the light is such as to make Poissonian variations negligible. Clearly, when h is above approximately 100 km, the low light level causes some statistical scatter to the plotted points. The points were read directly from the chart with a time constant of 1 second.

The values of $(I_0 - 0.9I_{10}')$ for $h > 80$ km lie on a curve which falls off much more slowly than Hunten's. The curve on this occasion steadily decreases to a value 3.5 arbitrary units, and then remains fairly constant. This value may then be taken as the nightglow intensity. It must be subtracted from the values of $(I_0' - 0.9I_{10}')$ in order to obtain the true twilight intensities; these are also plotted in Figure 1.

The twilight points still fall off more slowly than Hunten's curve. The discrepancy may be due in part to the neglect of multiple scattering in Hunten's theory, as a somewhat similar effect is observed in measurements of twilight-sky brightness [Hulburt, 1953]. Part, however, is also due to the poor correction for continuum resulting from our assumption that x is a constant. It has been pointed out [Hunten, 1955] that the scattered light becomes redder later in twilight, and hence, in our notation, x should be larger. Further measurements must be made to determine the effect of the reddening on the corrected curve in Figure 1.

From Figure 1 we find that the ratio (P) of

TABLE 1. Characteristics of Twilight Sodium Emission

Date		Z_c , km	P
Antarctic measurements			
2/5/59	evening	90	
2/6/59	evening	105	
2/20/59	evening	90	
Melbourne measurements			
6/17/59	morning	91	30
6/17/59	evening	95	...
6/21/59	evening	87.5	18
6/27/59	evening	92	...
6/28/59	evening	85	50
7/2/59	morning	95	15
7/7/59	morning	88	30
7/8/59	evening	91	48
7/9/59	morning	89.5	42
7/9/59	evening	92	47

twilight to nightglow intensity was about 47 for the evening of July 9, 1959. It is difficult to estimate the absolute value of P , which will be affected both by variation of x and by difference of the ratio D_2/D_1 in the twilight and the nightglow. Nevertheless, the accuracy of relative values of P should be correct to about 10 per cent.

In Table 1 are gathered estimates of Z_c for three evenings in the Antarctic and nine mornings and evenings in Melbourne; several values of P are also listed. In the Antarctic measurements, h was always less than 120 km, so that nightglow could not be isolated; the lack of two values of P in the Melbourne measurements was due to obscuring clouds in late twilight.

Discussion. The estimates of Z_c in the Antarctic are probably correct only to about 10 km. Those in Melbourne have a mean value of 90 km, which is slightly higher than Hunten's value of (85 ± 5) km and Blamont's value [Blamont and others, 1958] of 88 km. Estimates of Z_c derived from Hunten's curve may be in error because of variation in ozone content of the atmosphere [Hunten, 1956]. The average ozone content in Melbourne in June and July is about 0.30 cm, which is higher than that used by Hunten. Measurements of the ozone content in Melbourne on the days listed in Table 1 show no marked correlation with either Z_c or P , with the limited number of observations made. Uncertainty in estimates of Z_c arising from this

and other features probably limits the accuracy of our measured mean Z_c to (90 ± 5) km.

The values of P , the ratio of twilight to nightglow intensities, lie between 15 and 50. They show no correlation with Z_c . It appears that at least some of the variation of P is real, but more observations will be made to examine this point in detail. It is worth noting that a combination of Hunten's twilight data [Hunten, 1956] and Roach's nightglow measurements [Roach, 1955] for mean seasonal intensities would give P about 11.7 in summer and 14.3 in winter. It has been suggested by Roach [1956], however, that OH contamination may have contributed about 60 rayleighs to his nightglow intensities.² If so, the new values for P would be 18 in winter and 80 in summer. Clearly, more observations are required in both hemispheres.

A further interesting feature of the sodium D emissions is the variation of their intensity with time in the southern hemisphere. We have used an activated zinc sulfide screen as a constant light source, and have made preliminary measurements of nightglow intensity compared with this source.

We found that the mean summer intensity in Melbourne was about half the mean winter intensity. Further, the variations from night to night were less than 50 per cent about the mean value, and the variations during any one night

were less than 20 per cent about the mean.

Although these measurements would seem to indicate that variations of nightglow intensity in Melbourne are much less than those in the northern hemisphere, the results should be treated with caution until more detailed and systematic measurements are completed.

Acknowledgments. The author gratefully acknowledges the helpful discussions and assistance given by Dr. F. Jacka, Chief Physicist, and Mr. G. Cresswell, of the Antarctic Division. He thanks Dr. J. P. Funk, of the C. S. I. R. O. Division of Meteorological Physics, for providing data on the ozone content of the atmosphere.

REFERENCES

- Blamont, J. E., T. M. Donahue, and W. Weber, *Ann. géophys.*, **14**, 282, 1958.
 Hulburt, E. O., *J. Opt. Soc. Am.*, **43**, 113, 1953.
 Hunten, D. M., *J. Atmospheric and Terrest. Phys.*, **5**, 44, 1954.
 Hunten, D. M., *Sci. Rept. AR-21*, University of Saskatchewan, 1955.
 Hunten, D. M., The airglow and the aurorae, *Proc. Belfast Conf.*, edited by A. Dalgarno and E. Armstrong, Pergamon Press, pp. 114 and 183, 1956.
 Mayaud, P. N., and R. Robley, *Ann. géophys.*, **10**, 258, 1954.
 O'Brien, B. J., and G. de la Harpe, *J. Sci. Instr.*, in press.
 Roach, F. E., *Ann. géophys.*, **11**, 209, 1955.
 Roach, F. E., *Natl. Bur. Standards Rept. 5006*, 1956.

(Manuscript received August 18, 1959; revised October 5, 1959.)

² The existence of this reference was kindly indicated by the editorial referee.

On the Rotation of the Polar Ionospheric Regions¹

C. O. HINES

Radio Physics Laboratory

Defence Research Telecommunications Establishment

Defence Research Board, Ottawa, Canada

Abstract. The possibility of magnetic coupling between the polar regions of the earth's ionosphere and the interplanetary gas has led to the suggestion that the polar ionosphere may not rotate with the earth. The depth to which the effects of the interplanetary drag might penetrate is examined here with the aid of two simple models. The results are not conclusive, but they do indicate that heights as low as the *E* region may be involved.

Introduction. It has been suggested elsewhere [Hines, 1959] that the polar regions of the earth's ionosphere may not rotate with the earth, and that various fluid dynamic and electrodynamical consequences peculiar to the polar regions may thereby result. The operative mechanism is thought to be a drag effect exerted by the interplanetary gas and transmitted by magnetic viscosity through the exosphere to the lower ionospheric regions. This mechanism is identical with its main features to that discussed by Dungey [1958] for rotating stars, and suggested by him as a possible cause of the reduced speed of rotation observed at high latitudes on the sun.

The implications in the case of the terrestrial ionosphere are severe, but extremely difficult to analyze in any detail. It seems appropriate, therefore, to attempt a very simple preliminary calculation of the depth to which the drag of the interplanetary gas might be expected to penetrate, in order to gain some assurance in advance that the suggestion is indeed worthy of further attention. Two such calculations are outlined here. Though neither is at all conclusive, both suggest that a real problem does exist for ionospheric heights.

Linear motion. In the first instance, a model will be assumed in which a conducting fluid is the half-space $z \geq 0$, moves with speed V_0 in the $+x$ direction at the interface $z = 0$, and is held stationary at $z = \infty$; it is penetrated by a uniform magnetic field lying in the $-z$ direction whose induction is B_0 , it has Pederson

conductivity σ_1 and negligible Hall conductivity, and it has coefficient of viscosity η (all in mks units). The problem is to deduce its steady-state motion, $V = V(z)$, which will lie in the $+x$ direction.

The motion of the fluid will induce in it an electromotive force of magnitude VB_0 in the $+y$ direction. A current density of magnitude $\sigma_1 VB_0$ will then flow in that same direction, and it will interact with the magnetic field to produce a body force on the fluid with density $\sigma_1 VB_0^2$ in the $-x$ direction. The only other operative force is that of viscosity, and it has a density $\eta d^2V/dz^2$ acting in the $+x$ direction. The two forces must annul each other for equilibrium to be established, and so

$$\eta d^2V/dz^2 = \sigma_1 VB_0^2 \quad (1)$$

in the steady state. The solution of this equation, when coupled with the desired boundary conditions, leads to the expression

$$V = V_0 \exp -z/z_0 \quad (2)$$

with

$$z_0 = \eta^{1/2} \sigma_1^{-1/2} B_0^{-1} \quad (3)$$

for the fluid motion. Thus the impressed motion V_0 falls off exponentially in the $+z$ direction, with a 'velocity scale height' of z_0 .

Although the model just discussed bears little resemblance to the actual ionosphere, it may contain the features essential to a first understanding of the ionospheric problem. It is interesting to examine the order of magnitude of z_0 . Below the ionospheric regions conductivities are very small and z_0 is very large; a motion

¹This work was performed as part of project CC D48-95-11-01.

impressed from below may be expected to persist with little change. A typical Pederson conductivity for the E region would be 10^{-4} , however, and this combines with the representative values $\eta \sim 10^{-5}$ and $B_0 \sim 6.10^{-5}$ to give $z_0 \sim 5$ km. From these simple considerations, then, it seems that the atmospheric motion impressed from below may well be damped out low in the polar E region. (No corresponding suggestion is made for lower latitudes, since the field lines there are not connected through to the interplanetary gas; and, of course, such a suggestion would be incompatible with observation.)

Rotational motion. The second model to be discussed is identical to the first except that the interface $z = 0$ is now rotating about the z axis in such a manner that its vector velocity at any coordinate point $(x, y, 0)$ is given by

$$\mathbf{V}_0 = [-\omega_0 y, \omega_0 x, 0] \quad (4)$$

A steady-state solution might reasonably be sought, in which the fluid velocity at a coordinate point (x, y, z) would have the form

$$\mathbf{V} = [-\omega y, \omega x, 0] \quad (5)$$

with $\omega = \omega(z)$; the boundary conditions would then require $\omega(0) = \omega_0$ and $\omega(\infty) = 0$.

It might reasonably be supposed that, in these circumstances, the magnetic field lines might be 'wound up' much like a rope, being 'frozen in' to the fluid motion in part, and only 'diffusing through' the fluid when sufficient curvature of the field lines is attained—to use the jargon of hydromagnetics. In such a state, the magnetic induction might be expected to have the approximate form

$$\mathbf{B} = [-by, bx, -B_0] \quad (6)$$

with $b = b(z)$, which corresponds to spiraling lines. The associated current density would then be

$$\mathbf{J} = \mu^{-1} \text{curl } \mathbf{B} \\ = \mu^{-1}[-x db/dz, -y db/dz, 2b] \quad (7)$$

where μ is the permeability of the fluid. This current would be maintained not only by the electromotive force induced by the motion but also by some electrostatic field \mathbf{E} set up by space charge. The latter can be deduced from the relations

$$J_{x,y} = \sigma_1(\mathbf{E} + \mathbf{V} \times \mathbf{B}_0)_{x,y} \quad J_z = \sigma_0 E_z \quad (8)$$

where σ_0 is the longitudinal conductivity and second-order terms in b have been neglected. Insertion of (5) and (7) in (8), followed by application of the relation $0 = \text{curl } \mathbf{E}$, which must apply in a steady state, then leads to the conclusion

$$d^2 b/dz^2 = \mu \sigma_1 B_0 d\omega/dz \quad (9)$$

The current density (7), in conjunction with the magnetic induction (6), leads to a body force which lies only in the φ direction of a cylindrical coordinate system (r, φ, z) , provided that second-order terms in b are again neglected. The density of the force is $-\mu^{-1} r B_0 db/dz$. In the steady state, this force must be balanced by the viscous force, which is also confined to the φ direction and which has magnitude $\eta r d^2 \omega/dz^2$ in that direction; hence

$$\mu \eta d^2 \omega/dz^2 = B_0 db/dz \quad (10)$$

Equations 9 and 10 can now be combined and solved with the appropriate boundary conditions to yield

$$b/b_0 = \omega/\omega_0 = \exp -z/z_0 \quad (11)$$

with

$$b_0 = \mu \eta^{1/2} \sigma_1^{1/2} \omega_0 \quad (12)$$

and z_0 given by (3) as before. The neglect of second-order terms in b requires, at least for partial justification, that $r \ll B_0/b_0 \sim 10^4$ km in the E region, so the result is not prejudiced on this account in the polar regions. The earlier estimate of z_0 again appears pertinent, and with it its implications of damping in the E region.

A serious difficulty does arise, however. In the case of rotation, a radially outward centrifugal force density $\rho r \omega^2$ is introduced (where ρ is the fluid density). This can be balanced by a pressure gradient, but unless ω is independent of height the pressure then becomes a function of z : $p = p_0 \exp(r^2 \omega^2 \rho_0 / 2p_0)$ and $p/p_0 = \rho/\rho_0$ for an isothermal gas, with p_0 and ρ_0 constants. (Gravity further complicates this picture but does not affect its essential features.) As a consequence, vertical pressure gradients are brought into play, and these can exceed by a large factor the inductive forces that gave them birth. Alternatively, p can be maintained independent of height, but a radial force density of magnitude $\rho r(\omega^2 - \omega_0^2)$ must

en be taken into account. In neither case does an additional force act in the φ direction, which is the direction of prime importance in the determination of ω , so it would be difficult to conclude without further study that any drastic change must be made in the estimated damping of the rotational motion. Nevertheless, no firm assertion can be made of the form of that damping until a more self-consistent solution has been found. Attempts in this direction by the present author have so far ended in failure. It is not at all clear whether this failure results from a basically invalid assumption as to the form of the solution or from the neglect of some essential feature (such as boundaries in the x, y direction, with associated polarization charges) in the model adopted; both aspects of the analysis are suspect.

Conclusion. It seems possible at this stage

to draw only a very tentative conclusion: In the absence of inductive drag exerted through the exosphere by the interplanetary gas, the atmosphere could be expected to rotate with the earth at all heights; but in the presence of such a drag more complicated problems concerning the velocity distribution arise, and there are strong indications that these problems must be faced for heights as low as the E region in the polar ionosphere. Further study therefore seems to be warranted.

REFERENCES

- Dungey, J. W., *Cosmic Electrodynamics*, Cambridge University Press, 1958.
Hines, C. O., Motions in the ionosphere, *Proc. IRE*, 47, 176-186, 1959.

(Manuscript received October 9, 1959.)

Ion-Density Measurements in the Stratosphere

JOHN L. KROENING¹

*School of Physics
University of Minnesota
Minneapolis, Minnesota*

Abstract. The number density of small negative ions in the atmosphere was measured in 1958 on a series of eight balloon flights which reached altitudes of 115,000 feet over Minneapolis (45°N). A pronounced decrease in the number density of small negative ions was observed at the tropopause on seven of the eight flights, and similar decreases were observed at numerous temperature inversions in the stratosphere. The results strongly suggest the presence and bunching of 'dust' above the temperature inversions.

In the absence of dust and ionizing agents other than cosmic rays, the number density of negative ions increases with altitude to 50,000 feet, where a maximum of approximately 5500 ions/cm³ occurs. Above this altitude the number density decreases to a value near 2000 ions/cm³ at 115,000 feet; this result is contrary to the almost constant value expected in the stratosphere from cosmic-ray data and the Thomson theory.

Introduction. The ionization present throughout the part of the troposphere above the exchange layer, and the stratosphere, is generally thought of as arising from cosmic radiation alone, and its disappearance is attributed to volume recombination consistent with the Thomson theory [Loeb, 1939]. The experiment of Gish and Sherman [1936] (National Geographic Society balloon, Explorer II, 1935) indicated that the recombination coefficient for small ions is approximately proportional to the 1/3 power of the pressure and that Aitken nuclei are present in the stratosphere. However, the results obtained by Stergis and others [1955] and Woessner and others [1958] apparently did not indicate a disagreement with the Thomson theory nor show any evidence of nuclei at high altitudes. The measurements reported in this paper lend support to the observations made on Explorer II.

If quasi-stable conditions obtain and Aitken nuclei are absent, the equilibrium equation is $q = \alpha n^2$, where q is the ion production rate, α is the recombination coefficient, and n is the number density of small ions. Hence a knowledge of q and n is sufficient to determine α . In the range of altitudes to be considered q is the production rate of ions deduced from measurements of the intensity of cosmic radiation. Since

the variation of the intensity of cosmic radiation with height above the earth's surface is well understood, only the ion density needs to be measured to provide a check on the recombination coefficient.

Method of observation. In the experiments described herein plastic balloons were used to carry the instruments through the atmosphere. Eight flights took place during 1958. The instruments were suspended approximately 400 feet below the balloon as a precaution against the electrostatic effect of the balloon. The pressure, the ambient air temperature, and the number density of negative ions were measured on each flight. The pressure was monitored with two separate instruments which overlapped within the range of 80 to 10 mb. The air temperature was measured with the standard white-rod thermistor. The information was transmitted to a receiver by means of the standard radiosonde technique and on most of the flights was recorded on photographic film as well.

The ion collector. The ion-density measurements were made with a chamber consisting of two concentric aluminum cylinders between which an electric field is maintained to collect the ions as the air flows through the chamber. The geometry of the collector in conjunction with the electric field is adjusted so as to insure collection of all ions passing through the instrument. The fundamental difference between this

¹ Convair Fellow 1958-1959.

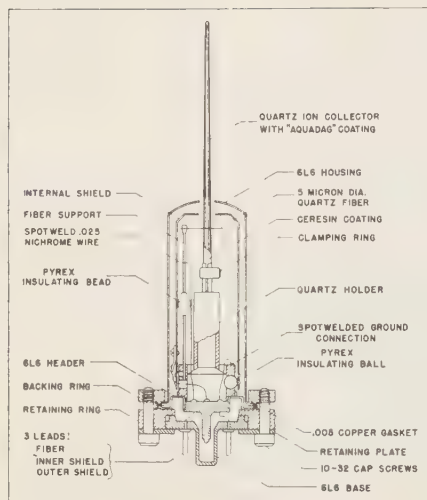


Fig. 1. The counting mechanism.

and a conductivity chamber is that the conductivity chamber operates in the ohmic region whereas the ion collector is designed to saturate.

The outside cylinder used in these experiments is a 6-inch length of tubing, 0.06 inches thick, having an outside diameter of 3 inches. The center rod is approximately 0.08 inches in diameter and 4.5 inches long. A wire mesh is placed across the open ends as shielding for the center rod.

The counting mechanism. The method used to count the ions as they are collected is similar to that developed by *Neher* [1953]. The counter (Fig. 1) is mounted in an aluminum shield which is welded to the side of the outer cylinder of the collector. The quartz rod of the counter emerges through a 0.2-inch diameter hole in the side of the outer cylinder so as to be the perpendicular bisector of the center rod of the collector. The center rod of the collector and the quartz rod of the counter are then clamped together so that the center rod is supported securely.

Near the support end of the quartz rod and close to its surface a gold-plated quartz fiber is mounted. This fiber is connected via its support to the positive terminal of a 270-volt battery, the negative terminal being connected to the outer cylinder of the collector. Assume initially that the quartz rod and the center rod

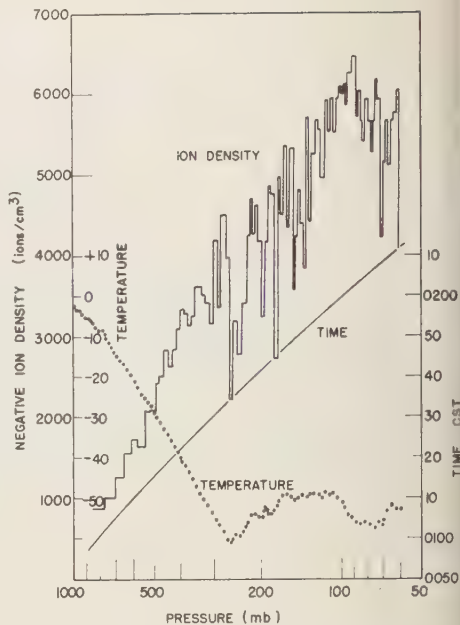


Fig. 2. The ion density, temperature, and time versus pressure on flight 1 (March 13, 1958).

of the collector are at the same potential as the fiber. As the rod assembly collects negative ions its potential with respect to the fiber changes. When the rod has collected a certain amount of charge the attractive force between the fiber and the quartz rod is large enough to pull the fiber against the rod, thus recharging it. Each charging of the rod can be used to generate a voltage pulse which is used in monitoring the time between adjacent pulses.

The charge per pulse deposited on the rod assembly is determined with an electrometer amplifier which drives a pen recorder. The uncertainty in the calibration and the inconsistency of the instrument are no more than a 3 per cent effect when taken together.

Results. Figures 2, 3, 4, 5, and 6 contain the launching dates and results of five of the eight flights. In each figure the time (CST), the temperature ($^{\circ}\text{C}$), and the ion density (ions/cm 3) versus the pressure in millibars for that flight are plotted. From the time and temperature versus pressure curves a rate of rise for the balloon is easily obtained. The ion density is then given by

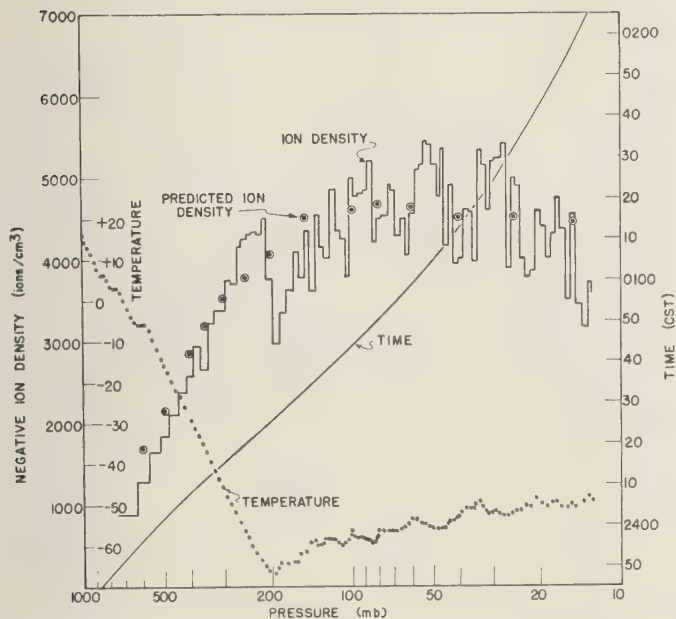


Fig. 3. The ion density, temperature, and time versus pressure on flight 2 (April 17, 1958).

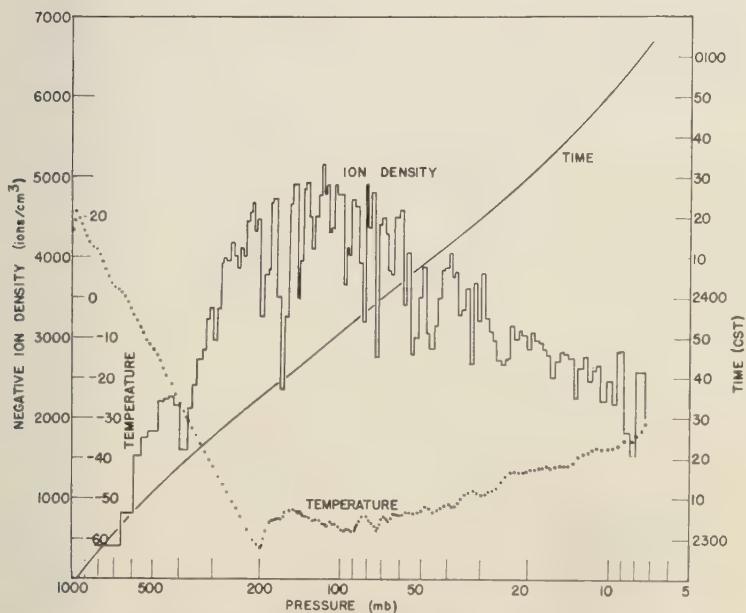


Fig. 4. The ion density, temperature, and time versus pressure on flight 3 (July 11, 1958).

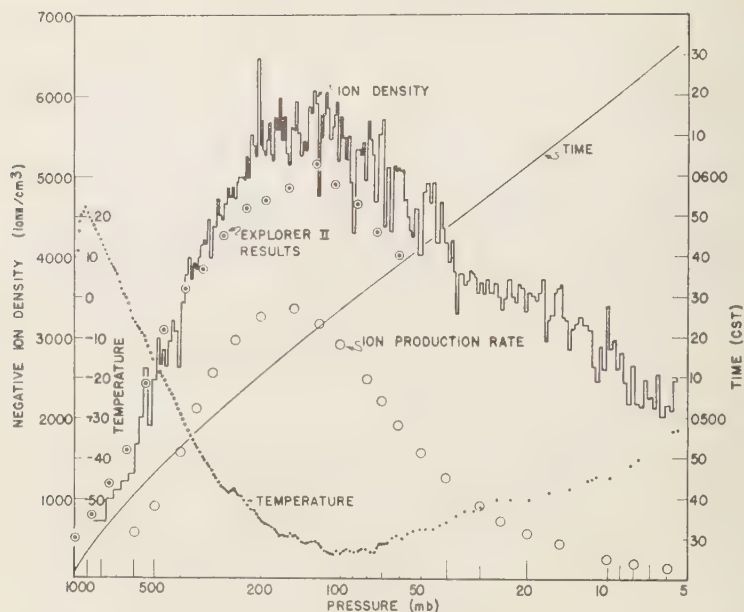


Fig. 5. The results of flight 4 (October 14, 1958) compared with the results of Explorer II; the scale of the production rate of small ions (ion pairs $\text{cm}^{-3} \text{sec}^{-1}$) is 1/100 of that for the ion density.

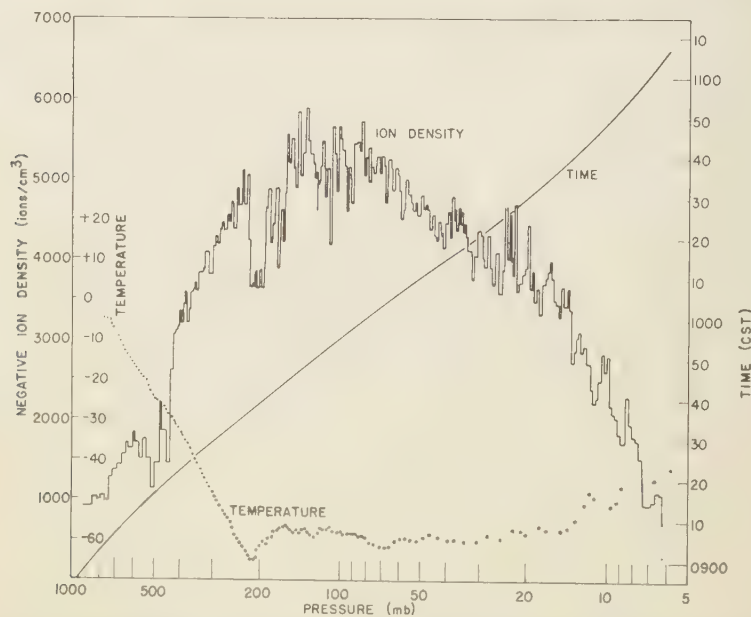


Fig. 6. The ion density, temperature, and time versus pressure on flight 5 (December 18, 1958).

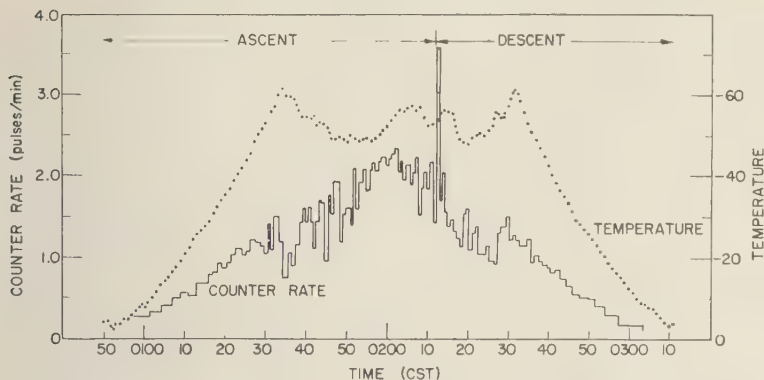


Fig. 7. The raw data (counter rate and temperature versus time) of flight 1.

$$n = RC/v\pi(a^2 - b^2)$$

here

R = counter rate in pulses/sec
 C = calibration factor (number of ions/pulse)
 v = vertical velocity of the balloon
 a = radius of the outer cylinder of the collector
 b = radius of the inner cylinder of the collector

Flights 1, 2, 3, and 4 were launched during clear sky conditions, whereas flight 5 was launched under a 50 per cent cloud coverage. The apparent decrease in the ion density over the pressure range of 600 to 420 mb in flight 5 is attributable to clouds, since the balloon was observed with a theodolite as it gradually disappeared into a very diffuse cloud at 600 mb. Flights 1 and 2 are peculiar in that they occurred during visible auroras.

The flights were conducted during the night hours, or else a sunshade was used because of the effect of photoelectron ejection from the aluminum collector. The threshold for this process occurs for incident radiation of wavelengths 2900-4100 angstrom units, depending on the condition of the aluminum surface. The effect of sunlight on the collector was observed on the first two flights (not included in this report) of this series. The occurrence of this phenomenon is evidenced by a large counting rate of the ion counter when the balloon is floating at ceiling altitude. Conversely, a negligible counting rate at ceiling indicates the absence of this effect and the proper functioning

of the instrument. This observation, however, suggests the possibility of using a similar type of instrument, with proper filters to select a suitable wavelength interval somewhere within the Hartley bands, for the detection of ozone.

During the other flight (not included in this report) the balloon failed at 100 mb and showed results similar to those of flight 3.

There are two points to be made regarding the data:

1. The variation of the negative-ion density with the temperature lapse rate, with special attention given to the effect at the tropopause in flights 1, 2, 3, and 5 and the absence of any such large effect on flight 4.

2. The general decrease of the ion density with pressure below 100 mb indicated on flights 3, 4, and 5.

Discussion. The apparent decrease in the ion density at the tropopause was observed on seven of the eight flights, the exception being flight 4. The question immediately arises whether the observations indicate instrumental difficulties or are truly representative of ambient conditions. At present, it appears that the only other variable that could result in the effect observed at the tropopause is the balloon velocity, which enters into the calculation of the ion density. This, however, does not seem to be the case because of the following considerations.

First, the time-pressure curves indicate no change in rate of ascent which would account for the decrease.

The second argument comes out of a more complete examination of the results of flight 1

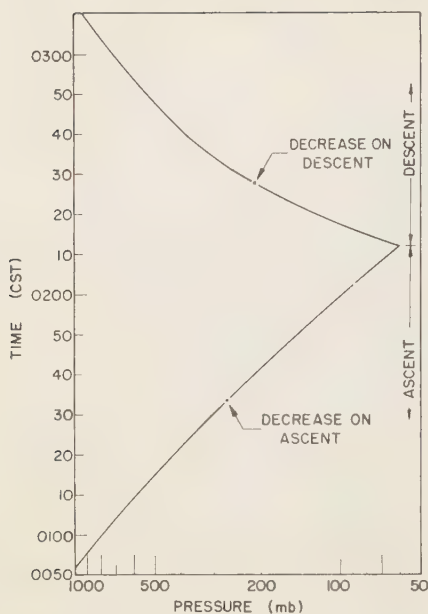


Fig. 8. Time-pressure curve for flight 1.

(Fig. 7), in which the counter rate and the temperature are plotted as functions of time. The time-pressure curve is shown in Figure 8. The balloon reached an altitude corresponding to 60 mb, at which altitude the equipment was separated from the balloon and descended, at first very rapidly, by parachute. The initial rapid descent did not allow the instrument to saturate; however, this has no effect on the argument.

A comparison of the ascent with the descent shows that the major part of the decrease on the descent occurred at a small inversion some 5000 feet above the tropopause, whereas on the ascent it was observed at the tropopause. Now the aerodynamic drag on the parachute is a smoothly varying function of the air density and the speed of descent and consequently cannot account for the result observed. Furthermore, the effect is not peculiar to every temperature inversion, as evidenced by its absence at the tropopause on the descent. From these arguments it is concluded that the data represent ambient conditions.

The existence of such large variations in the ion density must clearly require the presence of

Aitken nuclei. In addition to these experiments, and that of Explorer II, the measurements of *Bigg* [1956] on the detection of dust by twilight scattering and those of *Gergen* [1957] on the existence of anomalous high clouds indicate that perhaps not infrequently nuclei are present in the upper troposphere and the stratosphere. A mechanism for maintaining dust at such altitudes has been pointed out by *Durst* [1933, 1935].

If the measurements discussed in this paper and those referred to are related, it is possible that the dust observed affects the heat transfer process of the atmosphere. In fact, the numerous observations by *Gergen* show almost complete attenuation of the upwardly directed infrared radiation at levels in the vicinity of the tropopause, yet the Weather Bureau reported no clouds at this level on these occasions.

The number density of nuclei was not computed from these measurements, since the quantitative significance of such calculations is doubtful. However, a lower limit of the concentration of these nuclei is easily obtained from the graphs.

The absence of any large decrease at the tropopause on flight 4 is interesting when the temperature profile for this flight is compared with those having a decrease. The results suggest that an abrupt change in the lapse rate is needed to cause bunching of the Aitken nuclei above the inversions. On this assumption, the data of flight 4 may indicate that the nuclei are distributed in a more or less uniform manner throughout such a great range of altitude that they have no noticeable effect on the number density of small ions. The bunching hypothesis is most conspicuously demonstrated in flights 1, 2, and 3, where numerous small temperature inversions above the tropopause are strongly correlated with decreases in the ion density which are comparable to and even greater in magnitude than those that occur at the tropopause.

We turn now to a discussion of the general decrease of the ion density with pressure below 100 mb. Flights 3, 4, and 5 show this tendency, with a maximum occurring somewhere between 150 and 100 mb. If the assumption is made that the number density of Aitken nuclei is negligible except for the bunching mentioned above,

the general decrease is to be attributed to equilibrium between the cosmic-ray production rate and the volume recombination of small ions. In contrast to this, the ion density predicted from the cosmic-ray data and the Thomson theory is very nearly constant with altitude above 45,000 feet. In these three experiments the volume recombination is approximately proportional to the $2/5$ power of the pressure. Although this is not definitive, it serves to illustrate the disagreement with theory.

On the other hand, flight 2 shows no definite increase at higher altitudes, and, even though it occurred during a visible aurora, simultaneous measurements of the intensity of cosmic radiation indicated no extra ionization resulting from X-rays as was observed by Winckler and others [1958]. The compatibility of the predicted results with those observed on this flight is indicated in Figure 3.

The ion density derived from the conductivity measurements of Explorer II [Gish, 1939] is given in Figure 5 in order that a comparison may be made. The agreement is generally very good, but perhaps this is to be expected, since the measurements took place at the same latitude and at very nearly the same phase of the solar cycle upon which the intensity of cosmic radiation depends. Also plotted in this figure is the production rate of small ions for flight 4 as determined from Winckler's data (unpublished, available at IGY Cosmic Ray Data Center A, University of Minnesota, Minneapolis 14, Minnesota). Because of the constancy of the intensity of cosmic radiation over the year, except for periods of unusual solar activity, this curve suffices in the analysis of these experiments.

These observations are being continued with a slight change in the method of sampling. A positive displacement pump is to be used to eliminate the velocity of the balloon from the calculation of the ion density. This will allow the continuous monitoring of the ion density while the balloon is floating at ceiling altitude. Diurnal and seasonal variations and also those arising from changes in the production rate of small ions can be studied.

Summary. The following conclusions seem evident at the present time.

1. Aitken nuclei are present at higher alti-

tudes and tend to collect above temperature inversions, with the largest effect observed at the tropopause.

2. On repeated flights, the number density of small negative ions decreased with pressure below 100 mb; this result is contrary to the almost constant value expected in the stratosphere from the cosmic-ray data and the Thomson theory.

Acknowledgments. The author is grateful to all those who have assisted in this program. I especially wish to thank E. P. Ney, J. R. Winckler, and H. T. Mantis for their assistance throughout the experiments. I am indebted to R. Maas, J. Stoddart, W. Huch, and R. Howard for their proficiency in solving telemetering problems and carrying out the actual flight work. Finally, I wish to thank R. Hoffman, whose careful work permitted the accurate calibration of each unit.

This work was made possible by contract Nonr-710(22) jointly sponsored by the Office of Naval Research and the Signal Corps of the U. S. Army.

REFERENCES

- Bigg, E. K., Detection of dust and temperature inversions by twilight scattering, *Nature*, **177**, 77, 1956.
- Durst, C. S., Dust in the atmosphere, *Quart. J. Roy. Meteorol. Soc.*, **59**, 125-130, 1933, and *Quart. J. Roy. Meteorol. Soc.*, **61**, 81-89, 1935.
- Gergen, J. L., Atmospheric infrared radiation over Minneapolis to 30 millibars, *J. Meteorol.*, **14**, 495-504, 1957.
- Gish, O. H., *Terrestrial Magnetism and Electricity*, J. A. Fleming, ed., McGraw-Hill, New York, 149-230, 1939.
- Gish, O. H., and K. L. Sherman, The electrical conductivity of air to an altitude of 22 kilometers, *National Geographic Society, Stratosphere Series* **2**, 94-116, 1936.
- Loeb, L. B., *Fundamental Processes of Electrical Discharge in Gases*, John Wiley & Sons, New York, chap. 2, 1939.
- Neher, H. V., An automatic ionization chamber, *Rev. Sci. Instr.*, **24**, 99-102, 1953.
- Stergis, C. G., S. C. Coroniti, and A. Nazarek, Conductivity measurements in the stratosphere, *Proc. Conf. Atmospheric Elec., Air Force Cambridge Research Center, Geophys. Research Paper* **42**, 43-52, 1955.
- Winckler, J. R., L. Peterson, R. Arnoldy, and R. Hoffman, X-rays from visible aurorae at Minneapolis, *Phys. Rev.*, **110**, 1221-1231, 1958.
- Woessner, R. H., W. E. Cobb, and R. Gunn, Simultaneous measurements of positive and negative light ion conductivities to 26 km, *J. Geophys. Research*, **63**, 171-180, 1958.

(Manuscript received June 19, 1959; revised November 4, 1959.)

Atmospheric Tides and Ionospheric Electrodynamics

MARVIN L. WHITE

*Central Radio Propagation Laboratory
National Bureau of Standards
Boulder, Colorado*

Abstract. A brief review is given of the resonance theory of atmospheric tidal oscillations including both thermal and gravitational excitation. Semiempirical wind patterns for various latitudes are given in Figure 1, and the semiempirical variation of solar semidiurnal wind velocities with height are given in Figure 2, based on (1) solar barometric variations at ground level [Chapman, 1951], (2) wind measurements at balloon heights [Johnson, 1955], (3) radio meteor echo experiments [Greenhow and Neufeld, 1955], (4) E-region winds from radio pulse techniques [Briggs and Spencer, 1954], and (5) airglow cell movements [Roach and others, 1958]. Finally, current work on the extension of modern tidal theory (valid for a neutral envelope) into the dynamo region is discussed [White, in press].

1. INTRODUCTION

A most compelling feature of the earth's atmosphere is that, regardless of our everyday familiarity with the apparent vagaries and prices of ordinary weather conditions, certain atmospheric winds are world-wide. I am referring to the so-called tidal winds, which are both solar thermal and gravitational in origin. (There is a much smaller lunar-induced atmospheric wind.) These world-wide lunar and solar wind patterns may be predicted by a simple vector equation, 2, applicable from the earth's surface up to 100 km or more (see the semiempirical wind patterns in Figure 1 and their semiempirical height variations in Figure 2).

Above 100 km, that is, in the dynamo region of the ionosphere, the tidal theory for a neutral atmosphere is no longer entirely adequate. In particular, winds of the electron gas may depart considerably from that of the neutral gas. An extension of the tidal theory of Taylor [1936], Pekeris [1937], Weekes and Wilkes [1947], Sen and White [1955], and Siebert [1955] to include the dynamo region is reviewed here. Pertinent to this problem is the motion of a charged fluid plasma in the presence of a steady magnetic field which has been analyzed by Chapman and Cowling [1952]. The effects of diffusion upon the electron density are not included for the present, but variations due to thermal and tidal transport are, as are also the diurnal variations in the solar ionizing radiation. Inclusion of the diurnal

variations requires extension of the Chapman theory of layer formation in the ionosphere.

Although final numerical results are not yet available for the dynamo region, it is still possible to make contact with observation. The theory predicts a nonlinear electrodynamic viscosity and the production of higher harmonics (in time) in the dynamo region. Nonlinear effects are also predicted for the neutral envelope, due to Stokes viscosity; and, from the latter term, a 6-hour pressure variation with the correct phase and latitude dependence for the surface of the earth is predicted, in good agreement with the observation. For the ionosphere, some agreement between theory and observations exists, but results are not yet clear cut.

2. TIDAL THEORY

Modern tidal theory was fathered in 1936 and 1937 by G. I. Taylor [1936] and Chaim Pekeris [1937, 1939], respectively. Taylor showed that, under certain circumstances, the atmosphere might well have a whole series of resonant modes. Pekeris applied the theory to the then known model of atmospheric temperature distribution with height and found that, indeed, the atmosphere could have two important modes of vibration, one at 10.5 hours and another at 12 hours. The 10.5-hour period is precisely the period corresponding to the speed of a pressure wave sent out by the Krakatoa eruption of 1883. The 12-hour period is precisely the dominant com-

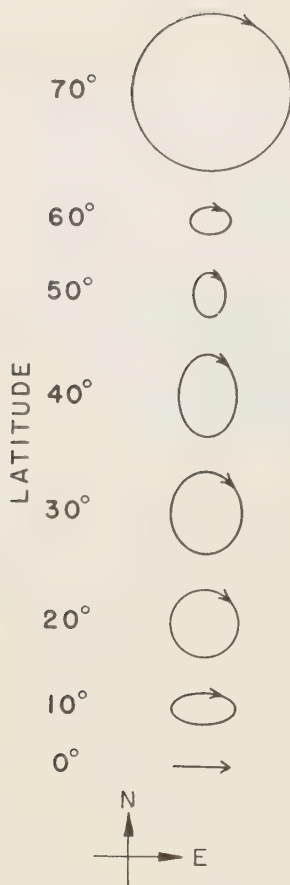


Fig. 1. Solar semidiurnal wind vector diagrams from equations 2 and 3 for $B = 0.14$. Scale values and phases given in Figures 2 and 7.

ponent observed in the solar barometric variation at the earth's surface.

Let us consider the origin of this 12-hour solar oscillation. If the oscillations are of thermal origin, the 24-hour barometric component should, of course, be greater than the 12-hour component, and if the solar-induced oscillations are gravitational in origin the solar tides should be half the corresponding lunar tides. Actually the solar is more important than the lunar oscillation by a factor of 15. In either event a way out of the dilemma is to suppose, along with Kelvin [Chapman, 1951], that the earth's atmosphere is resonant. If the solar-induced oscillations are strictly tidal, however, the atmosphere must be

highly resonant, for according to Hough [Chapman, 1951] the free period for atmospheric oscillations must differ from the imposed period of 12 solar hours by not more than 2 or 3 minutes since the lunar and solar semidiurnal periods differ by no more than 26 minutes. Chapman [1924] concluded after detailed mathematical development that solar thermal and solar tidal effects were roughly of equal importance.

Weekes and Wilkes [1947] extended the Pekeris [1937] treatment of gravitationally excited oscillations, and 4 years later Wilkes [1951] considered a thermal, instead of a gravitational, forcing function. The thermal source was supplied at the surface of the earth only. A further extension to include both a gravitational forcing function and thermal input at the earth's surface was subsequently considered by Sen and White [1955]. Among other things, they conclude that modern atmospheric profiles indicate the atmosphere to be only moderately resonant. On the one hand it is sufficiently resonant to discriminate against the 24-hour solar heating component in favor of the smaller 12-hour component; on the other hand, the resonance peak is so broad that the resonance amplification is roughly the same for the 12-solar-hour period and the 12-lunar-hour period. They conclude that the solar thermal oscillations must be more important than the solar gravitational oscillations by a factor of about 15.

The theory of the excitation of atmospheric oscillations has been extended by White [1956] to include thermal oscillations induced by heating at the surface, in the ozonosphere, and, in fact, throughout the atmosphere. A comparison with observations can be quickly made without going into elaborate programs of numerical integration. The theory shows that the thermal south-wind velocity component found from the resonance theory of tides can be expressed as

$$u = \frac{isg}{4a\omega^2} \frac{pH}{p_0} \frac{\left(\frac{d}{d\theta} + \frac{s \cot \theta}{f}\right) \Theta_{r,s}(\theta)}{(f^2 - \cos^2 \theta) \Theta_{r,s}(\theta)} \quad (1)$$

The ratio of the south-wind component to the east-wind component may easily be formed:

$$\frac{u}{v} = \frac{\left(\frac{d}{d\theta} + \frac{s \cot \theta}{f}\right) \Theta_{r,s}(\theta) e^{i\sigma t}}{i \left(\frac{\cos \theta}{f} \frac{d}{d\theta} + \frac{s}{\sin \theta}\right) \Theta_{r,s}(\theta) e^{i\sigma t}} \quad (2)$$

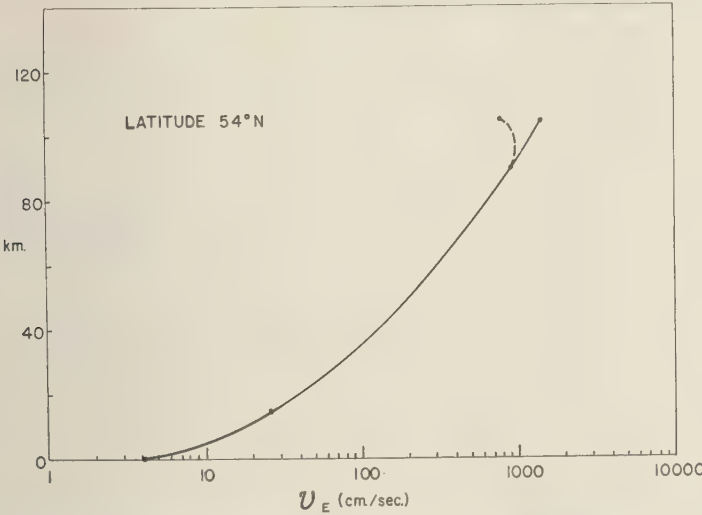


Fig. 2. Tentative semiempirical variation of the solar semidiurnal eastward velocity component with altitude for 54° N latitude. The solid line represents the wind for the neutral gas component; the dashed curve, for the electronic gas component.

where $i = \sqrt{-1}$, $\sigma = 2\pi/\text{period of oscillation}$, $\rho = 980 \text{ dynes/g}$, $a = \text{radius of earth}$, $\omega = \text{angular velocity of earth}$, $p = \text{pressure variation due to oscillation}$, $p_0 = \text{static pressure at given height}$, $f = \sigma/2\omega$, and $\theta = \text{colatitude}$; θ varies from 0° at the north geographic pole to 180° at the south geographic pole; for the solar semidiurnal oscillation,

$$\Theta_2^2(\theta) = P_2^2(\theta) - BP_4^2(\theta) \quad (3)$$

where $P_2^2(\theta)$ and $P_4^2(\theta)$ are associate Legendre functions and B is a constant. Equation 2 holds for both thermal and gravitational oscillations, and it shows that the south and east winds are in quadrature and that the wind vector diagram will form an elliptical Lissajous pattern in the east-south, east-west plane. It is possible to compare the resonance tidal theory with various observations. The observations consist of the solar semidiurnal winds observed (a) at ground level, discussed by Chapman [1951], (b) from balloon observations by Johnson [1955] at about 15 km, (c) by radio meteor echo technique discussed by Greenhow and Neufeld [1955] appropriate to the 90-km level, and (d) in the ionospheric region at about 105 km as compiled by Briggs and Spencer [1954].

Observed semidiurnal winds at ground level.

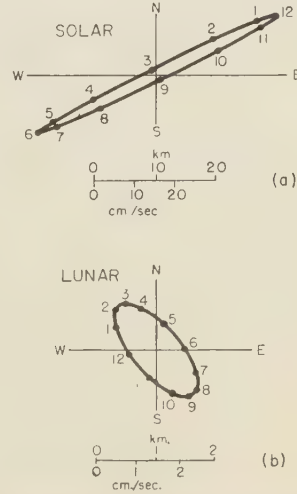


Fig. 3. Solar and lunar semidiurnal wind vector diagrams or actual displacement at ground level (island of Mauritius) following Chapman [1951]. Wind blows away from origin.

The solar semidiurnal wind vector diagram for Mauritius [Chapman, 1951] is given in Figure 3(a). It can be compared with wind data at the 15-km, 90-km, and 105-km levels in Figures 4,

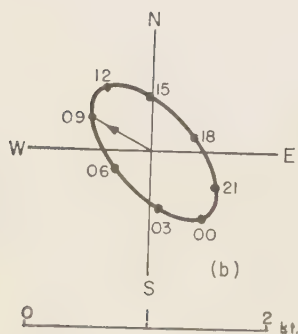
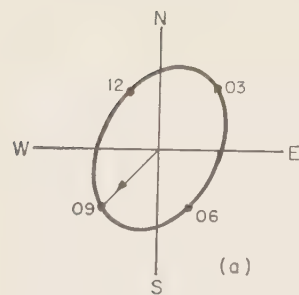


Fig. 4. Solar semidiurnal and diurnal wind vector diagram at balloon heights (≈ 15 km) following Johnson [1955]. The speed is given in knots.

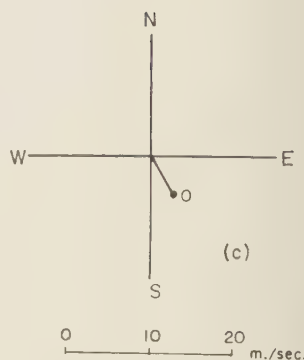
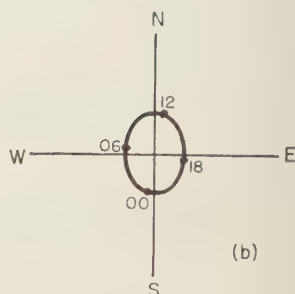
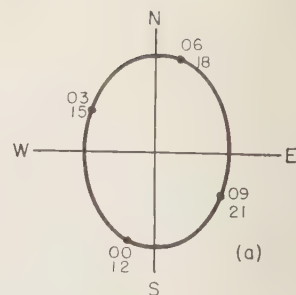


Fig. 5. Solar semidiurnal, diurnal, and steady wind components determined by radio meteor echo techniques at about the 90-km level following Greenhow and Neufeld [1955].

5, and 6, respectively. According to equation 2 the major and minor axes of the velocity ellipse should be aligned with the cardinal directions. Apparently the effects of surface friction are playing a role at Mauritius. Let us examine the sense of rotation. In the northern hemisphere, the rotation of the velocity ellipse is clockwise. For the southern hemisphere, it is counterclockwise (see Briggs and Spencer data for Lower Hutt). This is as expected theoretically from equation 3, with certain minor exceptions. For the diurnal variation, too, the rotation should be theoretically clockwise in the northern hemisphere, and this, again, is in agreement with observations at the 15-km and 90-km levels. Mauritius, being in the southern hemisphere, should have a counterclockwise wind for both lunar and solar semidiurnal winds according to equation 2. Figure 3b shows, unexpectedly, that the lunar wind is reversed; that is, it is clockwise.

It is not a difficult matter to compute what the Mauritius winds should be from equations 1

and 2, given the pressure variation p , at the ground (retaining the gravitational term omitted in equation 1). For p , the expression

$$p(W_2) = [P_2^2(\theta) - 0.14P_4^2(\theta)]$$

$$\cdot \cos(2t + 64^\circ)/4.05 \text{ mm} \quad (1)$$

is employed, where $p(W_2)$ is the pressure variation

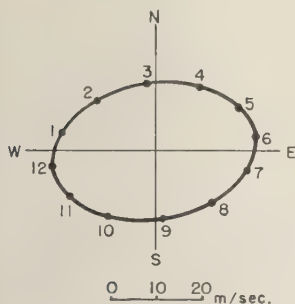


Fig. 6. Solar semidiurnal wind vector diagram of *E* region from combined Cambridge results [Biggs and Spencer, 1954] obtained by radio techniques.

corresponding to the traveling wave (a free translation of the German 'wandering wave') as opposed to the standing wave. The subscript 2 notes semidiurnal. For the present report, we will limit discussion to the traveling waves. The associated Legendre function used throughout will be defined according to

$$\int_{-1}^1 [P_n^2(\mu)]^2 d\mu = \frac{2}{2n+1} \frac{(n+2)!}{(n-2)!}$$

in accordance with usage of the NBS *Mathematical Tables for Associate Legendre Functions*.

It should be pointed out that equation 4 for (W_2) is at odds with the expression given by Schmidt [1928] and refined by Haurwitz [1956],

$$(W_2) = 35.51 \times 10^{-2} \text{ mb}$$

$$[P_2^2(\theta) - 0.047P_4^2(\theta)] \cos(2t + 68^\circ) \quad (5)$$

Equation 5 may not happen to agree with the reader's recollection of the Schmidt formula because it is usually expressed in the semi-normalized associate Legendre functions developed by Schmidt [Chapman and Bartels, 1940, p. 611] and used by Haurwitz [1956]. Equation 5 is an empirical expression based directly on world-wide solar semidiurnal pressure variations at the surface of the earth. Nevertheless, what follows on tidal winds, equation 4 will be used exclusively. It will in fact be demonstrated later that the coefficient $B = 0.14$ of equation 4 is consistent with world-wide wind data and equations 1 or 2, whereas equation 5 with a coefficient $B = 0.047$ is inconsistent. Thus the semiempirical Mauritius wind components are given by

$$\left. \begin{aligned} u \text{ (Mauritius)} &= 26.8 \text{ cm/sec} \cos 30^\circ (t + 5.2 \text{ hr}) \\ v \text{ (Mauritius)} &= 26.4 \text{ cm/sec} \cos 30^\circ (t + 2.2 \text{ hr}) \end{aligned} \right\} \quad (6)$$

from equations 1, 2, and 4. By south or east components we mean the direction toward which the wind blows, in contrast to the meteorological definition, which is just the reverse. The velocity components in equation 6 agree with the observations in order of magnitude. Comparison of phases will not be made. In a similar manner, the Mauritius lunar winds are computed from

$$p \simeq 55 \text{ microbars} \cos 30^\circ (t + 0.5 \text{ hr}) \quad (7)$$

from Chapman [1951], Figure 5, page 516, namely,

$$\left. \begin{aligned} u &= 0.95 \text{ cm/sec} \cos 30^\circ (t + 3.6 \text{ hr}) \\ v &= 0.93 \text{ cm/sec} \cos 30^\circ (t + 0.6 \text{ hr}) \end{aligned} \right\} \quad (8)$$

The magnitude of the winds is in agreement with Chapman's Mauritius results.

If the lunar ellipse is rotated through an acute angle of 50° so that the major axis coincides with the north-south direction, the theoretical

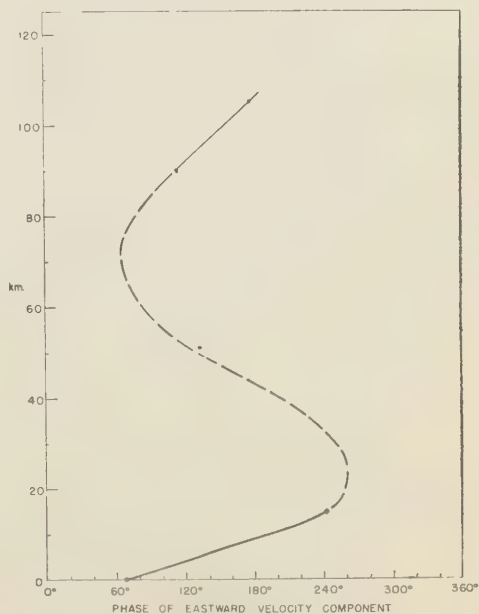


Fig. 7. Tentative semiempirical variation in phase of the solar semidiurnal eastward velocity component, v , as a function of altitude, where $v \propto \cos(30^\circ t + \text{phase in degrees})$ and where t is in hours (valid for either hemisphere).

and observed phase of u are in agreement and the phases for the east-west winds have been reversed. One final point: whereas all observed ellipses are found slightly rotated so that the major axes lie in the first and third quadrants, the opposite is true of the Mauritius lunar data. The two pieces of information in this paragraph and the reversed sense of rotation can be accounted for, interestingly enough, if the lunar winds have been reflected from a plane directed more or less in the north-south direction. However, one must also explain why the much larger solar winds are not similarly affected. The final solution must await further observational and theoretical developments.

Equation 6 will be considered the most trustworthy for the solar semidiurnal wind at the surface. To convert from 20°S latitude of Mauritius to the mean latitude of the British Isles (54°N) for comparison with other wind data,

$$\frac{u(\theta_1)}{u(\theta_2)} = \frac{\left[\left(\frac{d}{d\theta} + \frac{s \cot \theta}{f} \right) \Theta_r^s(\theta) \right]_{\theta_1} \left[\sin^2 \theta \right]_{\theta_2}}{\left[\left(\frac{d}{d\theta} + \frac{s \cot \theta}{f} \right) \Theta_r^s(\theta) \right]_{\theta_2} \left[\sin^2 \theta \right]_{\theta_1}} \quad (9)$$

where $f = 1$ for the solar semidiurnal oscillation while $s = r = 2$, and

$$\Theta_2^2(\theta) = [P_2^2(\theta) - 0.14P_4^2(\theta)] \quad (10)$$

thus, from equations 9, 10, and 6,

$$v(36^\circ) = 3.93 \text{ cm/sec} \cos 30^\circ (t + 2.2 \text{ hr}) \quad (11)$$

Equation 11 is used in establishing the ground point for Figures 2 and 7.

Winds aloft. Wind observations at the 15-km level, given by Johnson [1955], are

$$\left. \begin{aligned} u(36^\circ) &= 34.5 \text{ cm/sec} \cos 30^\circ (t + 4.5 \text{ hr}) \\ v(36^\circ) &= 26.3 \text{ cm/sec} \cos 30^\circ (t - 4 \text{ hr}) \end{aligned} \right\} \quad (12)$$

where the mean latitude of the British Isles ($\approx 54^\circ$) is taken. Johnson has used the meteorologist's definition of wind velocity in contrast to the definition used in equation 12. Equations 2 and 3 may be solved for B , thus

If the ratio $u(36^\circ)/v(36^\circ)$ from equation 12 is substituted into equation 13, $B(15 \text{ km}) =$ in close agreement with equation 10.

Similarly from the observations by Green and Neufeld [1955] at Jodrell Bank (90-km level)

$$\left. \begin{aligned} u(36^\circ) &= 14.5 \text{ m/sec} \cos 30^\circ (t + 0.7 \text{ hr}) \\ v(36^\circ) &= 9.0 \text{ m/sec} \cos 30^\circ (t + 3.8 \text{ hr}) \end{aligned} \right\}$$

and the corresponding semiempirical B value $B(90 \text{ km}) = 0.13$, again in substantial agreement with the law (10).

Wind data for the E region as collected by Briggs and Spencer [1954] have been similarly analyzed by White [1956] (see Table 1). B values were obtained from observed (u/v) ratios at Ottawa, Cambridge, and Lower Hutt as well as by comparing u components among the three stations by means of equation 9. Again the average B value turns out to be 0.14, although the B value from (u/v) for the combined Cambridge results was a notable exception ($B = 0.0$). For solar semidiurnal winds. Another conclusion to be drawn from Table 1 (see $|\Delta\phi|$ values) is that the E -region semidiurnal winds and from observational results (12) and (14) is that the u and v components are almost always in the phase quadrature, as predicted by present tidal theory. $|\Delta\phi|$ values represent departures from phase quadrature.

The world-wide winds at the ground, 15-km level, and 90-km level are still consistent with the author's previous conclusion [White, 1956] concerning E -region winds that equation (11) ($B = 0.14$) is appropriate for the latitude distribution $\Theta_2^2(\theta)$ which corresponds to a tidal pressure variation of equation 4 and not equation 5, i.e., Schmidt's formula. According to Schmidt the correct B value actually should be 0.0 and not 0.14. This discrepancy seems real. A quantitative explanation would be highly desirable.

The predicted ratios (u/v) for the semidiurnal velocity ellipses are given in Table 2 for the northern hemisphere; the table repeats its results for $90^\circ \leq \theta \leq 180^\circ$ except for a sign reversal.

$$B = \frac{P_2^2 \left[2 \cot \theta - \left(\frac{iu}{v} \right) \frac{2}{\sin \theta} \right] + \frac{dP_2^2}{d\theta} \left[1 - \left(\frac{iu}{v} \right) \cos \theta \right]}{P_4^2 \left[2 \cot \theta - \left(\frac{iu}{v} \right) \frac{2}{\sin \theta} \right] + \frac{dP_4^2}{d\theta} \left[1 - \left(\frac{iu}{v} \right) \cos \theta \right]} \quad (13)$$

TABLE 1

for the E region: (1) Semiempirical B values as defined by equation 3 appear in columns headed by B , (2) phase discrepancy $|\Delta\phi|$ between observed and theoretically determined wind component ratios given by $\frac{u(\text{station 1})}{v(\text{station 1})}$ or $\frac{u(\text{station 1})}{u(\text{station 2})}$ appear in columns headed by $|\Delta\phi|$; italicized values are for lunar diurnal wind. Table is taken from *White* [1956].

Locality and Observer	Ottawa (45°N) (J. H. Chapman)		Cambridge (52°N) results combined		Lower Hutt (41°S) (Burt)	
	B	$ \Delta\phi $	B	$ \Delta\phi $	B	$ \Delta\phi $
Ottawa (45°N) (J. H. Chapman)	+0.12 <i>+0.10</i>	13° 19°				
Cambridge (52°N) results combined	+0.14 <i>+0.11</i>	2° 1°	+0.30 *	9° 5°†		
Lower Hutt (41°S) (Burt)	<i>+0.17</i> §	77°‡ §	<i>+0.12</i> §	21° §	<i>+0.07</i> §	24° §

Value not accurate because of loss of significant figures in calculations.

Observations due to G. J. Phillips.

Comparison for corresponding seasons not available, causing large discrepancy.

No lunar wind data available for Lower Hutt.

stations 2 and 10 are used for this purpose. The corresponding ellipses are depicted in Figure 1 for every 10° of latitude. Figure 1 is valid for all altitudes; i.e., it is height independent. Mathematically speaking, it is a consequence of the fact that, in the expressions for u and v , the variables are separable. This is observationally substantiated by the fact that latitude dependence (10) seems valid for winds at all heights tested.

Of course, the scale for the ellipses will vary with height. To get some idea of the scale, we have plotted in Figure 2 the v components from stations 11, 12, and 14. For the 105-km level, we use the combined Cambridge results [Briggs and Spencer, 1954]

$$u(38^\circ) = 14 \text{ m/sec} \cos(2t + 78^\circ)$$

$$v(38^\circ) = 21 \text{ m/sec} \cos(2t + 177^\circ)$$

correcting to $\theta = 36^\circ$ via Table 2, second and third columns,

$$v(36^\circ) = 7.7 \text{ m/sec} \cos(2t + 177^\circ) \quad (15)$$

In comparison, for Ottawa observations [Briggs and Spencer, 1954] (at 105-km level)

$$u(45^\circ) = 34 \text{ m/sec} \cos(2t + 80^\circ)$$

correcting to $\theta = 36^\circ$ from Table 2, third column,

$$u(36^\circ) = 10.8 \text{ m/sec} \cos(2t + 80^\circ)$$

or from Table 2, second column,

$$v(36^\circ) = 5.29 \text{ m/sec} \cos(2t + 170^\circ) \quad (16)$$

Similarly from Ottawa E -region observations again,

$$v(45^\circ) = 23 \text{ m/sec} \cos(2t - 177^\circ)$$

or from Table 2, second and third columns,

$$v(36^\circ) = 4.65 \text{ m/sec} \cos(2t + 183^\circ) \quad (17)$$

There is a good deal of discrepancy in magnitude among equations 15, 16, and 17 in contrast to the over-all agreement in the B value for 105 km. Nevertheless, it can be concluded that the v values decrease above the 90-km level. The value 7.7 m/sec from equation 15 has been plotted in Figure 2. A possible explanation for this decrease may be found from the fact that the winds at 105 km correspond to irregularities of the electronic component as opposed to the neutral atmospheric component at 90 km and below. For the phase of the v component, there is much greater certainty in the E region [see White, 1956]. The phase of the v component, found from equations 11, 12, 14, and 15, is plotted in Figure 7. Figures 2 and 7, based as they are on information from only four heights, must be considered as only tentative, of course.

TABLE 2
(For $B = 0.14$)

θ , Colatitude, degrees	u/v	$u\alpha[(d/d\theta + 2 \cot \theta) \cdot (P_2^2 - 0.14P_4^2)]/\sin^2 \theta$
0	+1.0000	...
5	1.0011	-208.748
10	0.99931	67.329
15	0.99609	38.236
20	0.98498	22.092
25	+0.94928	-11.416
30	0.79729	3.7673
32	0.54021	1.3059
34	-2.2125	+0.87451
36	+2.0411	2.8024
38	1.3762	5.1578
40	1.4095	5.9836
45	1.3037	8.8500
50	1.2679	+10.639
55	1.2399	11.478
60	1.1991	11.475
65	1.1302	10.736
70	1.0183	9.3692
75	0.84990	+7.4879
80	0.61736	5.2140
85	0.32645	2.6740
90	0.0000	0.0000

Figures 1, 2, and 7 together with Table 2 form the basis for finding the semidiurnal winds at any time, latitude, and height at about 105 km and below.

Movement of airglow cells. Recent observations on what are termed airglow cells can give a possible clue to the motion of the neutral component (oxygen atoms) at the 100-km level. At a latitude of 40° , Roach and others [1958] obtain a wind, due south, at around 2100 hours. This is in agreement with the semidiurnal wind observations for Cambridge results combined. The rotation of the cells is clockwise, as would be expected for the northern hemisphere. The period of rotation has a mean value of about $5\frac{1}{2}$ hours according to Roach, instead of 12 hours; there is a larger spread in individual values, however. The mean speed for the cells is 93 m/sec; the corresponding semidiurnal wind speed for a latitude at 54° can be found as previously indicated, and when plotted in Figure 2 falls on the full curve. The conclusions to be drawn are that, whereas the horizontal component of electronic semidiurnal velocities

decreases with height owing to the magnetic field or electrodynamic viscosity (see the following section), the neutral semidiurnal wind velocity component in the horizontal plane continues to increase with height. This decrease of the electronic wind speed with height is consistent with observations made by Yerg [1956] at Puerto Rico. Working at two frequencies of 2.33 Mc/s and 4.57 Mc/s, he obtains wind speeds of 19 m/sec for the former and 12 m/sec for the latter so that, as the height increases, the electronic drift velocity decreases.

Comparison of airglow and Cambridge radio wind data. Let us examine the airglow data in more detail. In Figure 8, the 22 values of airglow winds from Roach and others [1958] are plotted in polar coordinates, showing both magnitude and direction. Only eight directions are specified as the original wind data are given only north, northwest, west, etc. Alongside each observational point is given the corresponding local time in hours and tenths of hours. It will now be of interest to compare this wind-velocity diagram for airglow with the semidiurnal wind at Cambridge (Fig. 6). The comparison will be limited to phases in time only. For convenience of comparison, the Cambridge local times associated with each of the eight directions given in Figure 8 as well as the Cambridge time interval associated with the eight directions. Also for convenience these time values have been recorded along the elliptically dotted path in Cambridge. We will first compute the probability that the local time for the 22 airglow wind values agrees with the local time interval for Cambridge. To find this probability, Figure 8 may be likened to a roulette wheel, the wind vector corresponding to the pointer and the eight possible time sections corresponding to the numbers on the roulette wheel. The probability of a given airglow observation's agreeing with the Cambridge results is therefore 1 in 8 in the case of pure chance. For 22 observations, there would be $22/8$ or about 3 coincidences. There are actually 5 coincidences. These values are in parentheses in Figure 8. Now, if we take into account the fact that the wind observations are at a height of roughly 105 km, while according to Roach and others the mean airglow height is lower (≈ 100 km), it can be seen from Figure 8 that the corresponding Cambridge times at 100 km can be obtained by increasing the given

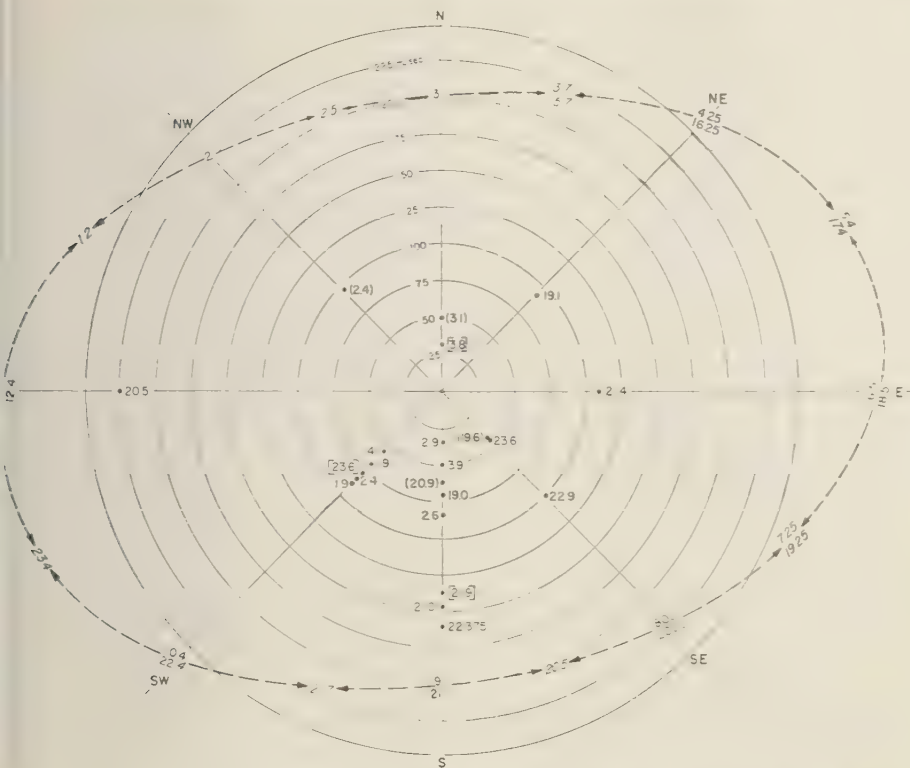


Fig. 8. Comparison of airglow wind data with the *E*-region semidiurnal winds of Figure 6. The elliptical path for the latter is represented by the dotted curve. Recorded values represent the time a wind blows in the direction indicated.

bridge times somewhat. The exact increase is uncertain, owing to uncertainties in the mean heights. However, for a time increase of 0.2–0.3 hr corresponding to a difference in height of 2.5 km, 8 out of the 22 observations agree with the Cambridge results instead of 5. The 3 values are in brackets in Figure 8. The probability of 8 or more successes in 22 tries, or a probability of $\frac{1}{8}$ for a given trial, is 0.04. As this probability is less than the conventional '1 in 20,' we may reject the hypothesis of random processes.

Comparison of airglow and Washington, D. C., radio wind data. The airglow wind data may be compared with daytime ionospheric radio measurements taken over Washington, D. C., for a 3-year period [Agy, 1957] (see Fig. 9). The daytime *E*-region winds obtained by

radio-fading techniques are known to be in close agreement with the Cambridge results combined. Now, however, we use *E*-region winds for September–December only, as airglow data are for October and November only. In this way the comparison will be more direct. Further, the latitude of Washington, D. C., is almost that of Boulder, where airglow observations were made. (Theoretically, a difference in latitude should not affect the phase.) Unfortunately, however, we are now comparing daytime with night-time data so that known 24-hour winds will affect the night airglow and daytime radio winds differently. Again, we must make a height correction. It will be in the same direction as before but larger; this conclusion was reached after an investigation of the median $h'E$ values obtained for Slough, Lindau/Hartz, and Wash-

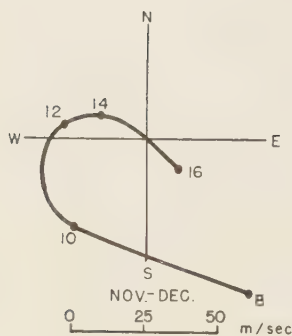
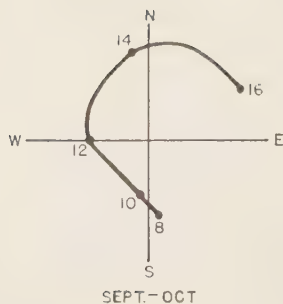


Fig. 9. Daytime *E*-region winds for Washington, D. C., by radio pulse techniques following *Agy* [1957].

ington, D. C. Slough and Lindau/Hartz are at approximately the same geographic and geomagnetic latitudes as Cambridge. For the comparison, data were obtained from the CRPL F Series for 1957 and 1958. The Slough values are recorded for every 5 km, so that comparison of Slough (March 1957–February 1958) with Washington, D. C. (September–December 1958), showed no significant difference. The Lindau/Hartz data (March 1957–February 1958), recorded every kilometer, gave an $h'E$ of 105 km compared with a Washington value of 110 km (September–December 1958). The Washington data are recorded to the nearest odd kilometer.

It is concluded then that, in comparing airglow and *E*-region winds at Washington, the correction will be in the same direction as before but more than twice that applied previously to Figure 8. When Washington and airglow winds are compared directly, 3 out of the 22 pieces of airglow data are in agreement; making the correction for a height difference amounting to 0.9 hour,

4 new points come into agreement, the 2 points remain, for a total of 7 of 22 points in agreement. The probability of this being chance is still only 1.5 parts in 100 as compared with the previous test of 4 parts in 100. Both tests, the airglow winds (representative of the neutral envelope) are about twice the velocity for the electronic component.

An explanation for the drift of airglow cells
A possible explanation for the drift of airglow will now be proposed: The movement of individual isophote lines across the circular isophote plots [as listed in *Roach and others*, 1958, Table 1] is explained, by and large, in terms of the semidiurnal and tidal oscillations in which a unit air mass experiences two perpendicular simple harmonic forces in the horizontal plane. For a latitude of 40°N , the paths of a given air mass will therefore describe a Lissajous pattern which is nearly circular in a clockwise direction. These winds, as we have seen, exist throughout the earth's atmosphere, *E* region included, the dominant component being semidiurnal owing to atmospheric resonance. The presence of semidiurnal tidal winds in airglow observations should therefore seem not unlikely, particularly when these wind velocities increase at a better than exponential rate with height. The apparent agreement of the 22 airglow observations with the extensive wind observations of radio pulse techniques is encouraging to present interpretation. Nevertheless, before any final conclusion can be reached many more airglow wind determinations will be required. Interestingly enough, the diameter of the nearly circular pattern mentioned above (corresponding to a maximum speed of 93 m/sec) will be half the mean diameter found for airglow cells.

A wind observation, in general, can be divided into two components, one turbulent or random and the other laminar (drift). Both components are of interest. By using enough data, the turbulent winds can be averaged out. For airglow winds the experimental time constant does not permit averaging. Regarding winds from visual meteorological observations, a cursory comparison of Figures 4 and 5 of *Liller and Whipple* [1954] with *Johnson* [1954], *Greenhow and Neufeld* [1955], and *Briggs and Spencer* [1954] suffices to show that the more instantaneous winds observed by Whipple and Liller are not directly comparable with the winds discussed here.

viscosity effects. In Figures 4, 5, and 6, wind or diagrams are reproduced from *Johnson* [15 km], *Greenhow and Neufeld* [1955] (105 km), and Cambridge results combined *Woods and Spencer*, 1954] (105 km), the last obtained by plotting u (38°) and v (38°) from station 15. At the 90-km level, the major and minor axis for both the diurnal and semidiurnal velocity ellipses are aligned with the cardinal directions. At the other levels studied, there appears to be a quite definite 'tilting' of the ellipses.

The tilt of the velocity ellipses in the E region must be accounted for by the collision of the ions (responsible for radio reflections) with overwhelmingly abundant neutral particles. This is the same collisional term appearing in the ionospheric theory. There, it is the polarization electric field that is rotated; here, it is the velocity vector. The similarity is of a basic nature. At the 15-km level and the ground level, Stokes viscosity or eddy viscosity will probably play a role.¹ *Chapman* [1924] found the effect of skin friction to be unimportant for the models examined. Qualitatively, however, we can see from Figures 4 and 10 that, since the wind velocity seems to increase at a greater than exponential rate with height and since the viscosity depends on the first derivative of the velocity, use of the present model may make viscosity effects appear more important. *Chapman* had assumed that the velocity increased only gradually with height. Thus, quoting from *Chapman's* 1924 paper, 'the variation of u with height will not produce any appreciable frictional effect on the relations near the ground.' In the light of Figure 2 let us reexamine the effect of surface friction by comparing the ratio of the acceleration term divided by the Stokes viscosity term, as follows:

$$\begin{aligned} \left| \frac{\bar{\alpha}_r \frac{\partial \mathbf{c}_r}{\partial t}}{\mu \nabla \cdot \nabla \mathbf{c}_r} \right| &= \left| \frac{i \sigma \mathbf{c}_r \bar{\alpha}_r}{\mu \nabla \cdot \nabla \mathbf{c}_r} \right| \\ &= \left| \frac{i \sigma \mathbf{c}_r \bar{\alpha}_r H^2}{\mu \mathbf{c}_r} \right| = \frac{H^2 \sigma}{\nu} \\ &= \frac{10^{10} \cdot 10^{-4}}{10^4} = 10^2 \end{aligned} \quad (18)$$

An apparent 'tilting' of the velocity ellipse can occur if the observations characterize a range of heights rather than a given height.

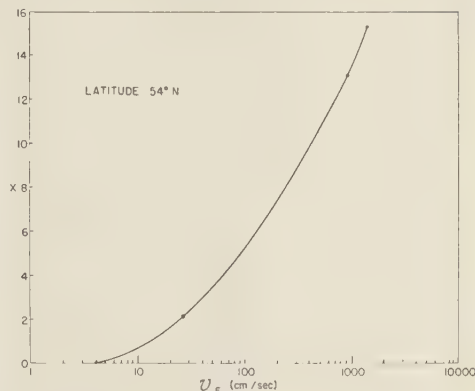


Fig. 10. The variation of v , eastward velocity component, for the neutral gas component with $x = \int dz/H$, where the scale height H is from the rocket panel model (recorded in Table 3); z = altitude.

where $\mathbf{c}_r \propto e^z$, say *Weekes and Wilkes* [1947], where $x = \int dz/H$, where

$$|\nabla \cdot \nabla \mathbf{c}_r| \propto |\mathbf{c}_r/H^2|$$

and where ν = kinematic viscosity = $\mu/\bar{\alpha}_r \simeq 10^4$ cm/sec for the ground [*Chapman*, 1924]. From equation 18, we find agreement with *Chapman*. From Figure 10, obtained from Table 3 and Figure 2, the increase is, however, at a greater than exponential rate, so that in reality

$$\mathbf{c}_r \propto \exp \left\{ \int \frac{dz}{F(H)} \right\}$$

where $F(H)$ is some function of the scale height H satisfying the inequality

$$\begin{aligned} \exp \left\{ \int \frac{dz}{H} \right\} &< \exp \left\{ \int \frac{dz}{F(H)} \right\} \\ &< \exp \left\{ \int \frac{dz}{\sqrt{H}} \right\} \end{aligned}$$

Let us now find the ratio A for the other extreme, $\mathbf{c}_r \propto \exp \left\{ \int dz/\sqrt{H} \right\}$. In this case

$$A \simeq \frac{i \sigma \mathbf{c}_r \bar{\alpha}_r H}{\mu \mathbf{c}_r} = \frac{\sigma H}{\nu} = \frac{10^{-4} \cdot 10^5}{10^4} = 10^{-3} \quad (19)$$

where now

$$|\nabla \cdot \nabla \mathbf{c}_r| \propto |\mathbf{c}_r/H|$$

where frictional effects are much more important than the acceleration term, an obvious impos-

TABLE 3
(For rocket panel model)

$z = \text{Height, km}$	$H, \text{Scale Height}$	$x = \int dz/H$
0.0	8.5240	0.0000
15	6.1455	2.0785
30	6.7870	4.4271
45	7.9410	6.4569
60	7.4050	8.3781
75	6.1280	10.6348
90	6.3560	13.0898
105	7.3960	15.2828
120	9.6660	17.0809

sibility. The true condition lies between the two extremes and opens the possibility that surface frictional effects may be important.

E-region lunar winds. Up to now we have discussed primarily solar-induced winds. Lunar semidiurnal *E*-region winds quoted by some observers are suspect. According to Table 2, lunar winds should be stronger in Washington, D. C., than in Ottawa and Cambridge. Yet, according to Greenstone as reported by *Briggs and Spencer* [1954], lunar semidiurnal winds could not be detected in the *E*-region over Washington. He concluded that lunar winds must be less than 10 m/sec. Preliminary calculations by the present author for the neutral envelope substantiate the Greenstone finding.

3. DYNAMO THEORY

Introduction. Electric currents above the earth's atmosphere were suspected more than a century ago when *Michael Faraday* [1832] suggested them as the cause of auroras. It was *Balfour Stewart* [1882], however, who, 50 years later, suggested that atmospheric currents cause the daily magnetic variations observed at the surface of the earth and advanced the dynamo theory. According to that theory, tidal oscillations of the conducting layers cutting across the permanent magnetic field induce earth-wide current systems in accordance with Faraday's law. The dynamo theory has since been elaborated by *Schuster and Chapman* [*Chapman and Bartels*, 1940] and others. In these works, atmospheric tidal velocities are introduced, usually with the assumption that the velocities are independent of height. The detailed varia-

tions of the current systems with height remain unknown.

An elementary and rapid presentation of dynamo theory can be made following *McL [Mitra, 1952]*. The same latitudinal representation of the semidiurnal tidal pressure variation on the ground as in the ionosphere is assumed

$$p = (AP_2^2 - BP_4^2) \sin [2(\lambda + t') + 154^\circ]$$

It is further assumed that

$$u = -\frac{d\phi}{a d\theta} \quad \text{and} \quad v = \frac{d\phi}{a \sin \theta d\lambda}$$

where a velocity potential ϕ has been introduced. When the pressure change p is small

$$d\phi/dt = -c^2(p/p_0)$$

or

$$\phi = -\frac{c^2}{p_0} (AP_2^2 - BP_4^2) \int \sin [2(\lambda + t') + 154^\circ] dt$$

where c is the velocity of sound. Substituting equation 20 for p into equation 23, ϕ is obtained from equation 21, the tidal velocities induced p , namely, u and v , are obtained; from

$$\mathbf{E}_s = \mathbf{v} \times \mathbf{B}_z \quad \text{and} \quad \mathbf{E}_w = \mathbf{u} \times \mathbf{B}_z$$

the south and west components of the electric field in volts per kilometer, say, can also be obtained.

Critique. Obviously, numerous important additions to the above approach to the dynamo problem are possible. As an example, Coriolis forces are capable of coupling north-south velocities into an east-west direction, and vice versa. It is simple to show by examination of the basic equations of motion that at certain latitudes Coriolis effects are important. As another example, it has already been pointed out in section 2 that the pair of constants A and B appearing in equation 20 should not be the same pair that appear in equation 23 for the velocity potential; further, the phase of the tidal winds changes more or less continuously with height (Fig. 7) so that the use of equation 20 is not justifiable.

equations 21 and 23 could be replaced by accurate expressions for the thermal wind velocities found from the resonance theory of (see equation 1). But, as we shall see, the field itself will be altered by the presence of charges.

Actually equations 21, 22, 23, 1, and 2, pertain to the neutral envelope. Above the 90-km level, as we have seen, the winds of the charged particles depart appreciably from the values expected for a neutral envelope. Apparently, the effect of the magnetic field on the charged particles is important even as low as the *E*-region, where a unified tidal-dynamo approach might be considered.

By a tidal-dynamo theory, we may mean many things, depending on how the tidal velocity is obtained. First, equation 20, obtained empirically from the earth's surface, could be assumed to hold for the upper atmosphere. For this case, the predicted velocities are off by a factor of 100.) Secondly, a wind field predicted by the theory of thermal and gravitational winds could be used to derive dynamo effects. This would be a great advantage over the dynamo theory for the earth's interior where the velocity field responsible for driving currents is unknown. A third possibility, used by White (in press), is to consider the actual changes in the wind field of the neutral particles as a background to their generating winds in the charged components. At first sight this effect might appear to be small, at least in the *E* region, where the charges are present only in tracer quantities. It is appropriate, however, to start integrations of the tidal problem at the 90-km level, where the degree of ionization is high, relatively speaking; as the integrations are continued down, the effect of the charged particles in modifying the wind field becomes cumulative. We shall presently show that, even changes in the neutral wind field were small, they would be essential to dynamo theory.

A new approach to the dynamo problem. The success of the resonance theory of atmospheric tides justifies its extension further into the ionosphere, into the plasma region. Considerable information is now available on the physics of fully ionized gas, that is, a two-component gas. Within the *E* and *F* regions, conditions are more complicated in that we are dealing with a three-component gas made up of neutral particles

and positively and negatively charged particles. Fortunately, however, the gas is only slightly ionized over that part of the atmosphere relevant to resonance tidal theory; in the *E* region, charged particles are present by at most 1 part in a million. Even in the *F* region, the neutral component is more abundant than the charged by a factor of 1000. Let us discuss the dynamics for such a medium.

Without the magnetic field, the movement of the three components will be indistinguishable and will be identical with the tidal oscillations in the gas as a whole. In the presence of the earth's dipole field, the conductivity becomes anisotropic and the mean motions of the three components will now differ from one another. Further, for the neutral component, one can more or less recover the basic equations of classical tidal theory, at least in form. A new term will arise representing the interaction of the neutral component with the charged components. This interaction or coupling term caused by collisional effects is analogous to an I^2R electrical power term, where the resistance R corresponds to a collisional frequency and the current I to the difference in mean velocities of the neutral and charged components.

Basic equations for resonance dynamo theory. We may now briefly summarize the basic equations for the neutral component or the gas as a whole: (a) the equation for hydrostatic equilibrium, (b) the perfect gas law, (c) the equation of continuity (conservation of mass), (d) the equation of energy conservation, (e) the vector equation of motion. The first three remain by and large as in classical tidal theory. The equation of energy conservation, (d), for the gas as a whole, however, now becomes [White, in press; 1957]

$$\begin{aligned} \bar{\alpha}_w (g C_w \bar{Q} / R) = & \bar{\alpha}_w \bar{p}_w \frac{\delta(1/\bar{\alpha}_w)}{\delta t} \\ & \begin{array}{cc} \text{thermal or} & \text{mechanical} \\ \text{radiant input} & \text{power} \end{array} \\ & + \bar{\alpha}_w C_w \frac{\delta \bar{T}_w}{\delta t} + \mathbf{j}_{\text{iond}} \cdot (\bar{\mathbf{E}}_w + \bar{\mathbf{c}}_w \times \mathbf{B}) \\ & \begin{array}{cc} \text{internal} & \text{electrical} \\ \text{power} & \text{power} \end{array} \\ & - \bar{\theta}_{ei} (\bar{\mathbf{C}}_e - \bar{\mathbf{C}}_i)^2 - \bar{\theta}_{in} (\bar{\mathbf{C}}_i - \bar{\mathbf{C}}_n)^2 \\ & \quad - \bar{\theta}_{en} (\bar{\mathbf{C}}_e - \bar{\mathbf{C}}_n)^2 \quad (25) \\ & \text{collision terms akin to a Joule's } I^2R \text{ term} \end{aligned}$$

where $\bar{n}_e \simeq \bar{n}_i$, while the vector equation of

motion for the gas as a whole becomes

$$\bar{\alpha}_w \frac{\partial \bar{\mathbf{c}}_w}{\partial t} + 2\bar{\alpha}_w \boldsymbol{\omega} \times \bar{\mathbf{c}}_w = -\bar{\alpha}_w \nabla \bar{\Omega} - \nabla \bar{p}_w$$

d'Alembertian for oscillatory motion	Coriolis force	external gravitational	total pressure gradient
--	-------------------	---------------------------	-------------------------------

$$-\bar{\alpha}_w \mathbf{g} + \bar{p}_w (\bar{\mathbf{E}} + \bar{\mathbf{c}}_w \times \mathbf{B}) + \bar{\mathbf{j}}_{\text{cond.}} \times \mathbf{B} \quad (26)$$

internal gravitational	electrostatic and Faraday forces	Faraday force	
---------------------------	--	---------------	--

The subscripts w , n , e , and i indicate gas as a whole, neutral, negative, and positive components. $\bar{\alpha}_w$ is the total mass density, and α_w the perturbed mass density. The symbol ρ is reserved for charge density. C_v is the specific heat at constant volume, R is the universal gas constant divided by the mean molecular weight, and $gC_v \bar{Q}/R$ is the quantity of heat per unit time per unit mass received by the gas through conduction, convection, or radiation. The bar over the symbol signifies a total value, i.e., the sum of the static and the time-varying components. $\bar{\mathbf{j}}_{\text{cond.}}$ is the charge flux vector, $\bar{\mathbf{E}}$ = the vector electric field strength, \mathbf{B} = vector strength of the earth's permanent magnetic field, $\bar{\mathbf{c}}_r$ = mean vector velocity of gas component $-r$, $\bar{\mathbf{c}}_r$ = mean peculiar velocity for particles of gas component $-r$, where

$$\bar{\mathbf{c}}_r = \bar{\mathbf{c}}_w + \bar{\mathbf{C}}_r$$

where $\bar{\mathbf{c}}_w$ = mass velocity of gas at any point,

$$\bar{\mathbf{c}}_w = \frac{1}{\bar{\alpha}_w} \sum_r \bar{\alpha}_r \bar{\mathbf{c}}_r \quad (27)$$

$\bar{\Omega}$ = tide producing potential, and

$$\bar{\theta}_{rn} \simeq n_r m_r \nu_{rn} \quad (r = e, i)$$

where n_r = number density, m_r = particle mass, ν_{rn} = collision frequency for neutrals and particles of gas component $-r$, where r refers either to electrons ($r = e$) or positive ions ($r = i$). Similarly

$$\bar{\theta}_{ie} \simeq n_e m_e \nu_{ie}$$

Finally, \bar{T} = total temperature and t is the independent variable 'time.'

To recover classical tidal theory in form, the I^2R term may, for example, be added to the heating term responsible for thermal oscillations. There is, however, one important difference. The appearance of the square in the I^2R term creates an important nonlinear effect, namely,

the effect of creating higher harmonics. In fact it turns out that the very essence of dynamo theory arises from nonlinear effects, a result new to current investigations [White, in press].

Now let us return to the motion of the charged particles. The presence of the magnetic field causes a $\bar{\mathbf{C}}_r \times \mathbf{B}$ term to appear in the equation of motion. The two oppositely charged components will tend to become separated, and drift current will be set up. Mechanical gravitational and thermal oscillations will be induced in the plasma as in the neutral component. The IR term also appears in the equations of motion for a given component. Now a term-by-term comparison shows that, in the present case of a slightly ionized gas, all the mechanical terms may be ignored compared with the electrical terms. Physically, this signifies that the oscillations set up in the neutral component are impressed upon the plasma through the collisional IR term are much more important than mechanical oscillations induced directly into the plasma. This simplification means that instead of a complex classical theory of tides for each of the three constituent gases, coupled through collisional terms, the equations of motion for the charged components each reduce to a pure electrical Kirchhoff's voltage law in the case of a slightly ionized gas, namely

$$\bar{p}_r (\bar{\mathbf{E}} + \bar{\mathbf{c}}_r \times \mathbf{B}) - \sum_s \bar{\theta}_{rs} (\bar{\mathbf{c}}_r - \bar{\mathbf{c}}_s) = 0$$

electrostatic and Faraday forces	\sum_s collision terms akin to an IR drop
--	--

$$r = e, i; s = n, e, i \quad (28)$$

As was mentioned previously in this section, wind-velocity field for the neutral component can be assumed, from which the current system can be derived. In present-day papers, the assumed model for the velocity field is somewhat arbitrary. In contrast to this method, the present author has considered the back action of the charged components on the dominant neutral component. This is not a pedantic point, but rather the essence of dynamo theory, for if the wind field remains unchanged all the components will be moving in the same direction. When the components move differentially, the wind field will change slightly from the field predicted by the tidal-resonance theory for a neutral envelope. It is this slight change that, though seemingly of second order, is responsible for tidal-dynamo

ry. In fact, the current density, ion drift velocities, and electric field are all expressible in terms of a small velocity difference between the gas moving as a whole and the velocity of the neutral component alone.

The differential equations of motion for the gas as a whole and the neutral component are almost identical but differ through a nonlinear term which arises out of the I^2R effect. In fact, the two equations are coupled through a nonlinear 'correction' term. The correction terms \bar{q}_w and \bar{q}_n (where w stands for gas as a whole and n for neutral component) are also expressible in terms of this velocity difference.

It may be worth while to indicate a few results of the new theory. Ordinarily, in tidal theory, one cannot get anything akin to a wave equation which includes time derivatives, for the following reason. The tidal wave equation is found by solving simultaneously 9 equations, some differential. Time derivatives are eliminated at the very beginning of the solution by assuming a time factor $e^{i\omega t}$. It is possible, however, to overcome the time dependency, which, with certain restrictions, turns out to be of third order in time. No second derivative in time is present. This is decidedly strange for a wave equation. Weekes and Wilkes [1947] have used results standard to electromagnetic wave theory in the boundary conditions for tidal resonance theory. It is now suggested that, though this analogy may be appropriate, further justifications are needed for. Supposing that the analogy is justified, the Weekes-Wilkes equation for the mean energy flow in the vertical direction, used to establish appropriate boundary conditions at the 'top' of the atmosphere, is in reality the oscillation of energy flow about a fixed height. In other words, the mean energy flow for the linear case will be zero.

Notwithstanding the above remarks, certain conclusions about boundary conditions at the top can be reached. Current observational evidence seems to substantiate the use of an atmospheric 'top' with a constant positive temperature gradient. Physically speaking the temperature rise in the F region and above would act as a barrier to oscillations originating from below so that the earth's atmosphere will behave as a cavity resonator with a moderate Q . For the modes of vibration for which this holds, the high ionization content beyond the barrier

will be unimportant, and present treatment, with the assumption of a slightly ionized gas throughout, is justified. Conversely, oscillations originating from above 400 km can easily penetrate, as such oscillations will 'see' a temperature decrease, at least down to the 80-100 km level. As a boundary condition for the atmospheric top, the present author would therefore suggest, in contrast with Weekes and Wilkes [1947], that at sufficiently high levels the energy flow be in a downward direction only.

Some further consequences of resonance dynamo theory. From the time-dependent wave equation the world-wide steady or seasonal winds, i.e., those winds that do not depend on the time, can immediately be obtained. The general circulation patterns (long-wave assumption included) can be given in terms of a local heating function $\bar{Q}_0(z)$ from White [in press],

$$\left. \begin{aligned} \bar{\psi}_0(z) &= -\frac{1}{\gamma H} \frac{\bar{Q}_0(z)}{1 - \gamma - \gamma(dH/dz)} \\ \text{where} \\ \bar{\chi}_0 &= (\bar{Q}_0/\gamma H) + \bar{\psi}_0(\theta, \phi, z, t) \end{aligned} \right\} \quad (29)$$

where ϕ = east longitude, z = altitude, γ = specific heat ratio, \bar{Q}_0 is the heat source function for the time independent case (hence subscript zero); $\bar{\chi}_0$ is the divergence of the velocities. The velocities themselves may be easily obtained from $\bar{\chi}_0$ using equations 11 and 12 of White [1956], which are valid for both thermally and gravitationally induced winds.

Another interesting study may be made from the time-dependent wave equation of motion (28), namely the matter of stability. The method consists in applying a small perturbation and examining the conditions for which it will grow or attenuate, attenuation representing a stable condition, growth an unstable one. The outcome depends, of course, on initial assumptions. In the case considered, the collisional effects (IR term) are expressed as the gradient of a scalar. It is unlikely, at least for the z direction, that collisional terms can be represented by the gradient of a scalar and included with the gravitational potential without losing the effect of friction altogether. We shall see that this is indeed the case. In the equation of energy conservation, the I^2R term is included, acting as a new source of energy. It can in fact be included in the heating or source term Q . It

has been found, interestingly enough, that a sufficient condition for stability is, in part, that the nonlinear I^2R term be variables separable. Actually the condition for 'neutral' stability is obtained; that is, the oscillations will neither grow nor decay, but remain at a constant amplitude. Apparently viscous effects had been neglected, after all, for the stability treatment.

Equations involving drift velocities and conduction currents \bar{j} can be greatly simplified in present theory by considering the actual conditions of temperature and density in the E and F layers. The final equation may still appear somewhat involved, but, for the inclusion of both E - and F -layer conditions, all the terms kept here are essential. For example [White, in press],

$$\bar{j} \cong \bar{\alpha}_w \nu_{in} \left[\Gamma^{-1} \beta_{en} (\bar{c}_w - \bar{c}_n)_{||} + \frac{\Gamma}{\beta_{en} B^2} (\bar{c}_w - \bar{c}_n) - \frac{1}{B^2} \mathbf{B} \times (\bar{c}_w - \bar{c}_n) \right] \quad (30)$$

where

$$\Gamma = 1 + \frac{\nu_{in}}{\nu_{en}} \quad \text{and} \quad \beta_{en} = \frac{\bar{p}_e}{\bar{\alpha}_e \nu_{en}}$$

Equation 30 shows that the conduction current depends not only on the collision of neutrals with electrons but also on the collision of neutrals with positive ions. Collisions of electrons and positive ions are apparently unimportant for determining \bar{j} , for these collisions were included in the initial equations.

Layer formation. A theory of the formation and properties of an ionized layer under certain simplifying conditions was given by Chapman [Mitra, 1952]. For a Chapman layer, the electron density at the height where the rate of ion production is a maximum is proportional to the square root of the cosine of the solar zenith angle. Equilibrium or noonday conditions are postulated. Strictly speaking, day-night effects are not considered. The E layer behaves roughly as a Chapman layer during the day. During the night, no new electrons are formed by solar ionizing radiation and there is a decay of charge density. The E layer, therefore, may represent a condition of strict solar control. In contrast,

ionization is present in the F region at all times and its behavior is frequently described 'distorted'; i.e., it behaves as an anomalous Chapman layer. It is thought that this distortion arises from the transport of electrons into a part of the layer in question, so that there is a competition between solar control and transport phenomena. The transport is governed in part by the to-and-fro oscillatory motions of the ions due primarily to earth-wide atmospheric thermal expansions and contractions and to a lesser degree by atmospheric tidal motions. The presence of the earth's magnetic field also serves to couple, for example, horizontal winds of the ionized components into a vertical direction, further serving to stretch the layers vertically and distort their shape. The charged components will tend to blow or move differentially, since opposite charges are oppositely bent by a magnetic field when collisional effects are ignored; thus, earth-wide current systems are set up (dynamo effect).

A unified theory for the time and space variations in ionospheric electron densities, that involves the theory of tidal and thermal oscillations, the dynamics of a plasma, and an expanded theory of layer formation. For the tidal-dynamo theory one obtains the two coupled differential wave equations for the gas as a whole and the neutral component; one also obtains a three-dimensional Kirchhoff's voltage law for each of the charged components. Since the charge density appears throughout, it is necessary to consider simultaneously layer-formation theory via the electron continuity equation [White, in press]

$$\frac{\partial \bar{n}_e}{\partial t} = \bar{q} - a_e n_e \nu - \bar{c}_e \cdot \nabla \bar{n}_e - \bar{n}_e \nabla \cdot \bar{c}_e$$

temporal change in number density	electron source by solar ionization	electron sink by recombination or attachment	transport terms
--	--	---	--------------------

(31)

The parameter ν can be either 1 (for attachment) or 2 (for recombination). The source function \bar{q} (or Q) is modeled according to

$$\bar{q}_r = \begin{cases} q_r(z) \cos \zeta & \text{days} \\ 0 & \text{nights} \end{cases} \quad (32)$$

where ζ = solar zenith angle. This model therefore supposes a Chapman layer to be a fair daytime representation, but includes the day-

effect and of course transport phenomena. At points, however, should be emphasized. At the present method is not sensitive to the assumption of a $\cos \zeta$ law. Equation 32 could be replaced by a square-wave assumption with the change; of course, the higher harmonics would then be slightly enhanced. The second point is that the transport term in equation 31 includes diffusion effects, as does the final equation, 33, below.

For the case $\nu = 1$, the semidiurnal variation of electron number density $\bar{n}_{e,2}$ for example, is given by [White, in press]

$$\bar{n}_{e,2} = \frac{a_1 - 2i\sigma_1}{a_1^2 + 4\sigma_1^2} \left[\bar{q}_2 - \bar{c}_{e,2} \cdot \left(\frac{1}{a_1} \nabla q_0 - \frac{q_0}{a_1^2} \nabla a_1 \right) - \frac{q_0}{a_1} \bar{\chi}_{e,2} \right] \quad (33)$$

The source functions \bar{q}_r are obtained from spherical harmonic analysis of equation 32. Similar results can be obtained for other harmonics and for the case $\nu = 2$. The transport term involving the divergence of the electron velocity $\bar{\chi}_{e,2}$ can be found by solving the above equation simultaneously with the differential equations of tidal motion, diffusion effects being neglected for the present. For the case $\nu = 2$ and the steady state or zeroth harmonic, the Chapman equation [Mitra, 1952] (equilibrium solution) is recovered

$$\bar{n}_{e,0} = (q_0/a_2)^{1/2} \quad (34)$$

Comparison of predicted nonlinear effects with observations. Even at the present stage of the theory, it is possible to make comparison with observation. For example, the Cavendish tables [Thomas and Robbins, 1955] of the time variations of electron densities at given true heights were harmonically analyzed. From Figure 11 there is a much larger ratio of second and third harmonic to the fundamental in equatorial regions than can be accounted for from the solar source function; from the spherical harmonic analysis of equation 32, the first harmonic could be twice the second for strict solar control. \bar{R} nonlinear effects seem to offer a reasonable explanation.

In the neutral envelope, too, one might expect small nonlinear effects arising from collision of neutrals with neutrals via a Stokes viscosity (of second degree in the equation of energy conservation). This now seems firmly established from our

examination of 'tidal' surface pressure variations harmonically analyzed by Kertz [1956]. The dominant term of the 6-hour pressure variation at the ground is explained as a self-coupling of the 12-hour pressure variation due to ordinary dynamic viscosity:

The 6-hour pressure wave consists of many components

$$W_4^4 = \sum_{\eta=4}^{\infty} T_{4,\eta}^4$$

where, following Kertz,

$$T_{\sigma,\eta}^{\sigma} = \text{constant } P_{\eta}^{\sigma}(\theta) \sin(\sigma t^* + s\phi + \epsilon) \quad (35)$$

where t^* = Greenwich time. The most important term is $T_{4,4}^4$ with $\epsilon = 225^\circ$. The corresponding h value is 15.3 km, which is apparently far removed from resonance conditions ($h \simeq 7.9$ km). It is difficult to explain this term on the basis of the so-called resonance theory unless there is a nonlinear coupling of the important $T_{2,2}^2$ term with itself, that is, a self-coupling of the 12-hour oscillation to produce the important 6-hour component.

Let us recall (see equation 5) that

$$T_{2,2}^2 \simeq \text{constant } P_2^2(\theta) \sin(2t + 158^\circ) \quad (36)$$

If

$$T_{4,4}^4 = \text{constant } [T_{2,2}^2]^2,$$

then, from equation 36,

$$T_{4,4}^4 \propto P_4^4(\theta) \sin(4t + 226^\circ)$$

in close agreement with the experimental value of Kertz (equation 35) and in support of conclusions stated above.

4. CONCLUSION

This report stresses world-wide semidiurnal winds from the ground level to the 105-km level. The dependence of these winds on latitude and height is shown in Figure 1 and 2, respectively; the variation of phase with height appears in Figure 7. The results are obtained semiempirically and are meant to be only tentative.

While the semidiurnal winds predominate over the steady and 24-hour component at the 90-km level and above (Figure 5), present tidal theory should allow us to predict the diurnal

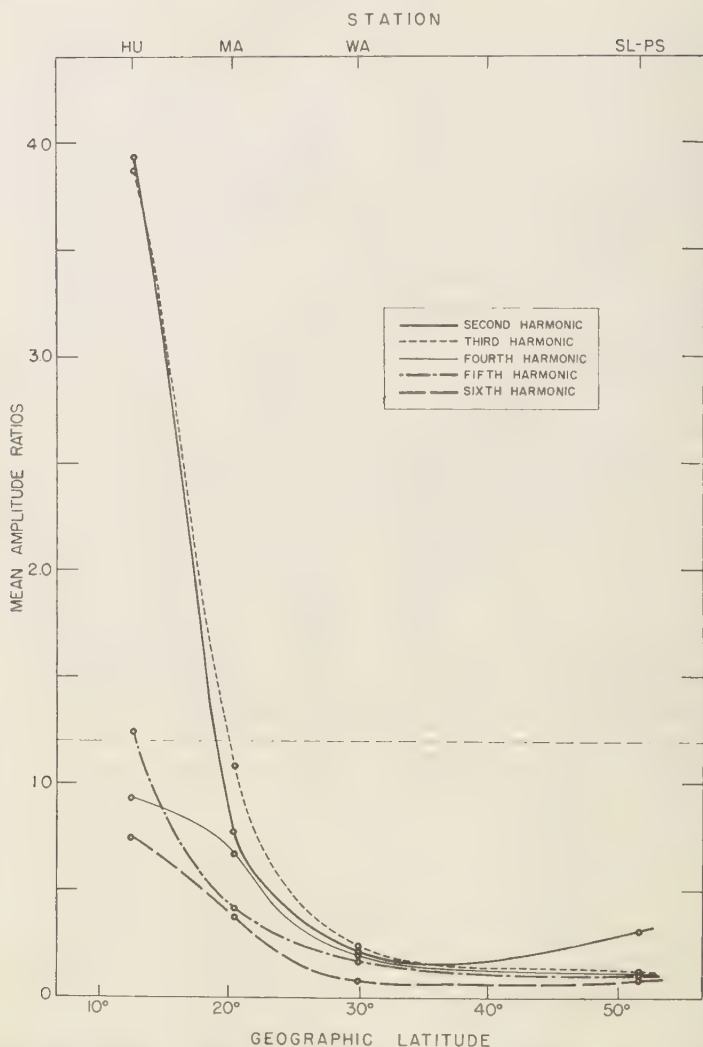


Fig. 11. Harmonic content of the daily variation in electron density (using Cavendish table [Thomas and Robbins, 1955] at a given height (280-300 km) at Huancayo (HU), Maui (MA), Watheroo (WA), Slough (SL), and Port Stanley (PS), expressed as ratios to the first harmonic. Weighted mean values for the first harmonic of the electron density in units of 10^5 cm^{-3} are 1.26185 (HU), 2.69474 (MA), 2.95497 (WA), 3.70578 (SL), and 3.25484 (PS).

and steady winds, the latter through equation 29, pertinent to the general circulation. By and large, the resonance theory of atmospheric tides seems capable of explaining the normal daily movements of the neutral envelope and even the electronic gas component. This can be seen, particularly, in the existence of a single B value

(see equation 2), namely 0.14, capable of characterizing all world-wide semidiurnal winds from the earth's surface through the 105-km level.

From semiempirical evidence, it is concluded that, in the E region, the semidiurnal wind for the electronic gas component is only half that of the neutral component (found from the

movement of airglow cells), emphasizing the importance of electrodynamic viscosity.

In section 3 a new approach to dynamo theory is obtained by extending the resonance theory of the ionosphere into the dynamo region. It appears that the electrified layer will oscillate nonlinearly through an I^2R Joule term which couples the neutral gas component to the electrified gas component in the ionosphere. The same theory also predicts nonlinear effects due to Stokes viscosity, and this effect is strongly substantiated by the presence of a 6-hour pressure variation at the surface of the earth with the appropriate phase (in time) and latitude dependence.

Finally, the resonance tidal theory plus its extension into the dynamo region forms the basis for a unified dynamical and electro-dynamical theory with the potential for determining normal world-wide daily variations in the important physical parameters from the earth's surface up to and beyond the F layer. The physical parameters are the number densities, temperatures, pressures, vector drift velocities for all constituents, the vector electric field, the vector current densities, and finally the scalar electron charge density responsible for ionospheric radio propagation.

REFERENCES

- Aggarwal, V., Results of ionospheric drift measurements in the United States, *Polar Atmospheric Symposium*, Part 2, Ionospheric Sections (Symposium at Oslo, July 1956), Pergamon Press, 1957.
- Biggs, B. H., and M. Spencer, Horizontal movements in the ionosphere, *The Physical Society, Repts. Progr. Phys.*, **17**, 245, 1954.
- Chapman, S., The semidiurnal oscillation of the atmosphere, *Quart. J. Roy. Meteorol. Soc.*, **50**, 165, 1924.
- Chapman, S., and J. Bartels, *Geomagnetism*, vol. II, Clarendon Press, Oxford, p. 768, 1940.
- Chapman, S., Atmospheric tides and oscillations, *Compendium of Meteorology*, edited by T. F. Malone, American Meteorological Society, p. 510, 1951.
- Chapman, S., and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases*, Cambridge University Press, chap. 18, 1952.
- Faraday, M., Experimental researches in electricity 2nd series 5. Terrestrial magneto-electric induction, *Phil. Trans. Roy. Soc. London*, **122**, 177, 1832.
- Greenhow, J. S., and E. L. Neufeld, Diurnal and seasonal wind variations in the upper atmosphere, *Phil. Mag.*, ser. 7, **46**, 549, 1955.
- Laurwitz, B., The geographical distribution of the solar semidiurnal pressure oscillation, *Meteorol. Papers*, **2**, (5), New York University Press, 1956.
- Johnson, D. H., Tidal oscillations of the lower stratosphere, *Quart. J. Roy. Meteorol. Soc.*, **81**, 1, 1955.
- Kertz, W., The seasonal variations of the six-hourly planetary pressure and temperature waves, *Sci. Rept. 5*, Research Div., College of Engineering, New York University, Project 299, October 1956.
- Liller, W., and F. L. Whipple, High-altitude winds by meteor-train photography, *Rocket Exploration of the Upper Atmosphere*, edited by R. L. F. Boyd and M. J. Seaton, Interscience Publishers, p. 112, 1954.
- Mitra, S. K., *The Upper Atmosphere*, The Asiatic Society, 1952.
- Pekeris, C. L., Atmospheric oscillations, *Proc. Roy. Soc. London, A*, **158**, 650, 1937.
- Perkeris, C. L., The propagation of a pulse in the atmosphere, *Proc. Roy. Soc. London, A*, **171**, 434, 1939.
- Roach, F. E., E. Tandberg-Hanssen, and L. R. McGill, Movement of airglow cells, *J. Atmospheric and Terrest. Phys.*, **13**, 122, 1958.
- Schmidt, A., *Handbuch der Experimentalphysik*, Geophysik, **25**, part 1, p. 172, 1928.
- Sen, H. K., and M. L. White, Thermal and gravitational excitation of atmospheric oscillations, *J. Geophys. Research*, **60**, 483, 1955; see also M. Siebert, On the theory of free and forced tidal-like oscillations in the earth's atmosphere, doctoral dissertation, University of Göttingen, 1955.
- Stewart, B., Terrestrial magnetism, *Encyclopædia Britannica*, 9th ed., p. 36, 1882.
- Taylor, G. I., The oscillations of the atmosphere, *Proc. Roy. Soc. London, A*, **156**, 318, 1936.
- Thomas, J. O., and A. R. Robbins, *Tables of F₂ Layer Electron Density on International Quiet Days*, ser. A, nos. 1 and 2, Cavendish Laboratory, Cambridge, England, June and October 1955.
- Weekes, K., and M. V. Wilkes, Atmospheric oscillations and the resonance theory, *Proc. Roy. Soc. London, A*, **192**, 80, 1947.
- White, M. L., Gravitational and thermal oscillations in the earth's upper atmosphere, *J. Geophys. Research*, **61**, 489, 1956.
- White, M. L., On the atmospheric dynamo theory, *J. Geophys. Research*, **62**, 329, 1957.
- White, M. L., Thermal and gravitational atmospheric oscillations—ionospheric dynamo effects included, *J. Atmospheric and Terrest. Phys.*, in press.
- Wilkes, M. V., The thermal excitation of atmospheric oscillations, *Proc. Roy. Soc. London, A*, **207**, 358, 1951.
- Yerg, D. G., Observations and analysis of ionospheric drift, *J. Atmospheric and Terrest. Phys.*, **8**, 247, 1956.

Some Observations of Ionospheric Faraday Rotation on 106.1 Mc/s¹

R. A. HILL AND R. B. DYCE

*Stanford Research Institute
Menlo Park, California*

Abstract. The polarization twist imposed on 106.1 Mc/s radio waves by the ionosphere has been investigated by using the moon as a passive reflector, the purpose being to determine the total electron column density even at altitudes above the known ionosphere. Because the antenna is capable of being continuously directed at the moon for 12 consecutive hours, observations are possible from the pre-dawn ionization minimum to the noon-time maximum. A true-height profile computed from vertical-incidence ionosonde data of September 16, 1957, suggests that the total electron content throughout the entire ionosphere is not a constant factor of the integrated electron content computed up to the level of maximum ionization density.

Faraday rotation. A VHF radio wave traveling through the ionosphere in the presence of a longitudinal component of the earth's magnetic field is broken into two counterrotating circular polarizations traveling at different phase velocities. When recombined, the linear resultant wave is found to be rotated (the Faraday effect) at its double passage through the ionosphere by an amount given by [Evans, 1957]:

$$= \frac{e^3}{4\pi^2 m^2 \epsilon_0 c^2 f^2} \cdot \int NB \cos \theta \sec X \, dh \quad \text{radians} \quad (1)$$

where e is the electronic charge, m the electron mass, f the wave frequency, N the electron density, B the earth's magnetic induction, θ the angle between the ray and magnetic field direction at ionospheric levels [Butler and Garriott, 1959], assuming a spherically stratified ionosphere, and X the angle between the ray and the zenith at the point where the ray passes through the ionosphere. Since $B \cos \theta \sec X$ varies by only 10 per cent over the ionospheric region, it is taken outside the integral and an average height is used for calculations. For a wave frequency of 106.1 Mc/s:

$$= 4.20 \times 10^{-12} B \cos \theta \sec X \cdot \int N \, dh \quad \text{radians} \quad (2)$$

$B \cos \theta \sec X$ may be calculated by assuming a dipole field for the earth in conjunction with the values of azimuth and elevation of the ray.

The technique was first applied by British workers [Evans, 1957] to determine the ionospheric electron content each moon transit. Ambiguity was resolved with the aid of two frequencies (about 1 per cent apart). A bistatic experiment using steerable antennas and a receiving parabolic dish equipped with a rotating feed have recently been used [Bauer and Daniels, 1958] to study the night-time ionosphere. Canadian workers [Blevins, 1957] have published preliminary results showing fluctuation in the amount of polarization twist occurring within about 2 minutes. The experiment to be described has the advantage of a steerable antenna capable of continuous observation over the entire period that the moon is above the horizon.

Equipment. Measurements of Faraday rotation for several days in September 1957 were taken using the Stanford Research Institute 60-foot parabolic reflector as the transmitting and receiving site and the moon as reflector (see Table 1). Technical limitations permit the use of only one wave frequency, and so the total electron content cannot be found unambiguously. However, changes of electron content can be measured readily by observing the rate at which the received power transfers from one polarization channel to the other.

Experimental observations. Figure 1 is a plot of the apparent number of electrons per square meter in a vertical column through the ionosphere as determined by Faraday rotation

¹ This work was supported by Rome Air Development Center under Contract AF 30 (602)-762.

TABLE 1. Characteristics of Radar Used for Lunar Observations

Location	37.4°N, 122.2°W
Frequency	106.1 Mc/s
Transmitter power	50 kilowatts (CW)
Antenna	60-ft parabolic dish (manually programmed to follow moon)
Antenna beamwidth and gain	12° (24 db gain)
Pulse width	2.5 _{sec}
Receiver bandwidth	~200 cps
Receiver sensitivity	~3 × 10 ⁻¹⁸ watt
Transmitted polarization	Vertical
Receiver polarization	Vertical and horizontal
Display	Dual pen recorder (Brush)

as a function of time (see equation 2). The zero for each plot is arbitrary because the minimum number of rotations is not certain. It is noted that the electron content begins to increase at about 0400 or 0500. Sunrise at the earth's surface for these days was about 0554 PST. Figure 2 is a plot of the electron content by Faraday rotation for September 16 in terms of electron density.

This curve should be compared with Figure 3, a plot of the computed number of electrons in a

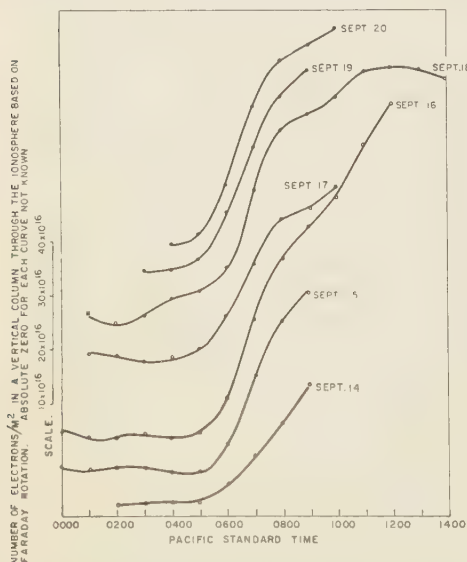


Fig. 1. Plot showing change of electron content with time for several days in September 1957.

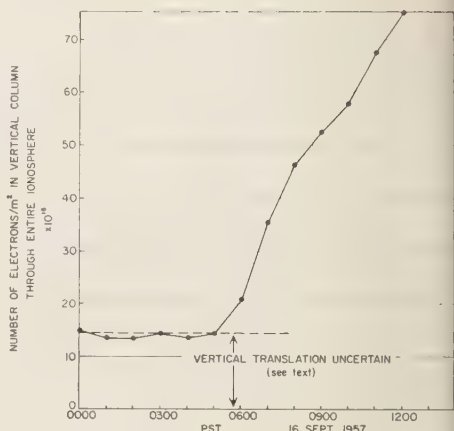


Fig. 2. Variation of the number of electrons per square meter in a vertical column through the ionosphere based on Faraday rotation data, September 16, 1957.

vertical column through the ionosphere up to the height of maximum electron density based on sweep-frequency ionosonde data. A parabolic fit was employed near the nose of the electron-density profile. The curves of Figures 2 and 3 indicate that the electron content begins to increase rapidly around 0500, the rate of increase tapering off around 0800 and increasing again after 0900. Since the tapering of the Faraday rotation curve (which depends on the entire ionosphere) is less than the tapering of the ionosphere sounding curve (which depends only on the lower ionosphere), it appears that the production of ions in the upper ionosphere continues at a rapid rate while that in the lower ionosphere temporarily suffers a decrease. At 1100 a similar effect occurs; namely, there is a continued production of ions in the region above the height

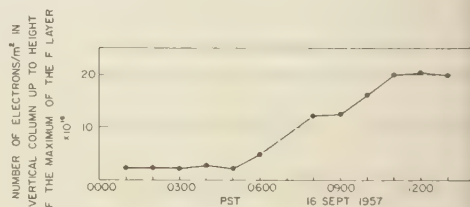


Fig. 3. Number of electrons per square meter in a vertical column from the ground up to the height of the maximum of the ionosphere, based on ionosonde data, September 16, 1957.

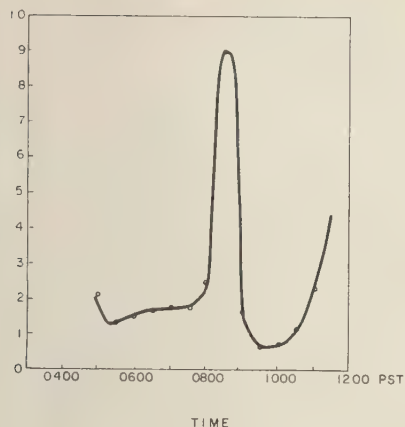


Fig. 4. Ratio of the rate of change of the total electron content to twice the lower ionosphere electron content as a function of time, September 19, 1957.

the peak of the F region, although the lower ionosphere has reached a maximum. In the notation of Bauer and Daniels:

$$\Delta n_f / \Delta t = k(\Delta n_p / \Delta t) \quad (3)$$

where
 $\Delta n_f / \Delta t$ = average rate of change of total electron content (slope of Fig. 2).
 $\Delta n_p / \Delta t$ = average rate of change of twice the lower part of the ionosphere (twice the slope of Fig. 3).
 k = proportionality factor.

then:

$$k = \frac{\Delta n_f}{\Delta n_p} \quad (4)$$

This ratio is plotted in Figure 4 as a function of time. According to observations by Evans [1957] and Bauer and Daniels [1958] the total electron content is twice that of a symmetrical parabolic model and remains fairly constant. This is greater than the ratio of about 1.5 expected for a 'Chapman' layer shape. On the basis of the data presented in Figures 2 and 3, the factor k appears to be far from constant. Such a procedure is, of course, dangerously dependent upon minor errors such as arise in the scaling of the ionosonde records.

The anomaly between 0800 and 0900 is

probably related to a daily nonuniform change of critical frequency observed at about 0800 or 0900 on every day in which data were taken. Furthermore, as is shown in Figure 5, the peak of the F region was also unusually low at this same time. Both these effects contribute to a smaller electron content or a decreased slope of Figure 3 in the interval 0800–0900. The peak value of k is primarily due to this anomaly.

The practice of using the maximum of the F layer as the dividing line between the known and the unknown is possibly unwise. A better choice in the examples shown might be to adhere to a constant height of, say, 300 km, integrating up to this level. Dangerous guessing would be required when the actual maximum lies below the 300-km level, and information is lost when the actual maximum lies above the 300-km level. The advantage of this more elaborate procedure would be to partially smooth the curve shown in Figure 3 and to tend to diminish the large peak shown in Figure 4.

An exact determination of the total electron content by the Faraday rotation technique is uncertain in this experiment because the remaining rotations in the early morning hours cannot be determined unambiguously. However, the data of Figures 2 and 3, when plotted against each other, approximate a straight line sufficiently accurately to allow extrapolation to the zero electron content point. This technique was used to determine the vertical translation in Figure 2. Once this assumption is made, the ratio of total electron content through the entire ionosphere to the electron content below the F peak can be determined as a function of time (Fig. 6). This plot, being the integral of Figure 4, is much smoother than Figure 4. By this technique it would appear that about twice as many electrons are above the F peak as are below it.

Conclusions. This research suggests that the number of electrons in the upper parts of the ionosphere exceeds the electron content derived by the assumption of a parabolic distribution. Apparently the ratio between the electron content of the upper portions and that of the lower portions does not remain constant as a function of time in the hours between sunrise and noon. The reason for this variability of the shape of the ionosphere is not yet known and would require further extension of theory or experiment.

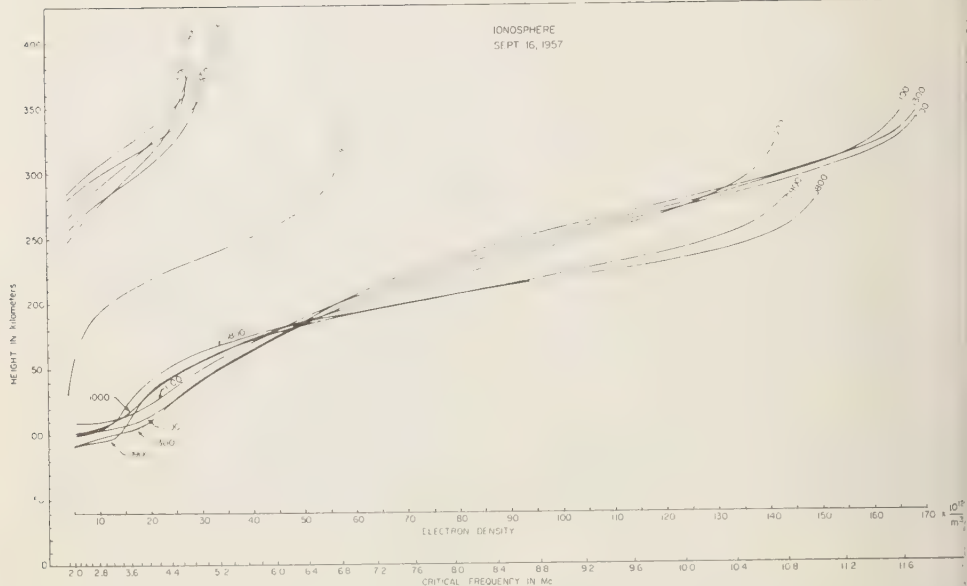


Fig. 5. Computed electron density profiles for September 16, 1957.

and consideration of lateral ionospheric gradients. It would be particularly interesting to examine the change of ionosphere shape with geomagnetic disturbance.

A second frequency in order to resolve ambiguity is an obvious extension of this experiment. Another improvement would be to preserve the phase between the two linear polarizations, in order to decide whether polarization is increasing or decreasing. The use of a system involving linear receiving antennas set at an angle other

than 90° has also been suggested as a means of determining the direction of rotation as a function of time, particularly at those times when the received polarization may occasionally be constant.

Acknowledgments. The task of maintaining and operating the radar required the efforts of many persons. The contributions of Lambert Dolphin, Myles Berg, Frank Firth, and Cecil Barnes are particularly acknowledged. Machine computations were greatly facilitated through the cooperation of Owen Garriott, of Stanford University.

REFERENCES

- Bauer, S. J., and F. B. Daniels, Ionospheric parameters deduced from the Faraday rotation of lunar radio reflections, *J. Geophys. Research*, **63** (2), June 1958.
 Blevins, B. C., Ionospheric studies by the lunar radar technique, *Nature*, **180**, 138-139, July 26, 1957.
 Butler, H. S., and O. K. Garriott, Magnetic field calculations, *Radio Propagation Lab. Rept. of Project Y/3243*, Stanford University, March 28, 1959.
 Evans, J. V., The electron content of the ionosphere, *J. Atmospheric and Terrest. Phys.*, **1**, 259-261, 1957.

(Manuscript received October 2, 1959; revised November 5, 1959.)

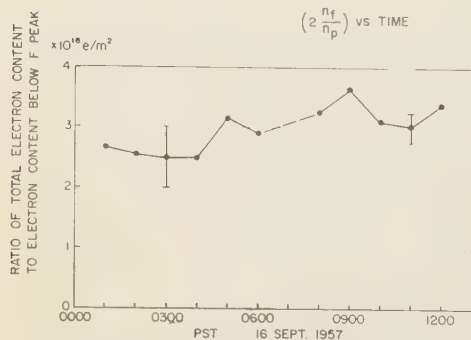


Fig. 6. Ratio of total electron content through the entire ionosphere to the electron content below the F peak.

The Lower Frequency Limits for *F*-Layer Radio Propagation

B. FULTON, O. SANDOZ, AND E. WARREN

*Defence Research Telecommunications Establishment
Ottawa, Canada*

Abstract. The band of frequencies propagated via the ionosphere by the high-angle ray and that propagated by the low-angle ray are both limited at their low-frequency ends by reflections that occur at lower ionospheric heights. Methods are developed in this paper for the calculation of these limits.

Introduction. Oblique incidence sounders are used to record, as a function of frequency, the time required for radio waves to travel by ionospheric reflection between distant locations. The records for the Ottawa-Slough (England) and the Ottawa-Saskatoon circuits show that there is a well defined lower limit to the band of frequencies propagated by the *F* layers in daytime. Separate and generally different lower frequency limits exist for the high-angle and low-angle modes of propagation (Figure 1). At frequencies below these limits, both these modes are obscured by reflections that occur in the lower regions of the same layer or in a lower layer. Methods of establishing the lowest frequencies propagated by a given mode will be discussed in this paper.

The conditions that determine these lower frequency limits may be inferred by using the concept of the reflection locus [Lejay and Lepechinsky, 1950; Petersen, 1951]. The reflection locus for the frequency *f* is a graph of correspond-

ing values of apparent height *h'* and elevation angle κ that satisfy the equation

$$f_o(h') \sec \phi(h', \kappa) = f = \text{constant} \quad (1)$$

This equation is the condition for reflection of the wave of frequency *f* in a vertical plane of the ionosphere. Here $f_o(h')$ is the frequency of the wave reflected at vertical incidence at height *h'*.

The high-angle ray. The lower frequency limit for propagation by the high-angle ray can be described by reference to the reflection loci in Figure 2. A similar figure will be used later to discuss the lower frequency limit of the low-angle ray. The reflection locus for waves transmitted from *T* has been drawn for each of the frequencies f_1, f_2, f_3 . The regions in the ionosphere from which waves of a particular frequency are reflected to the receiver *R* correspond to points that must lie not only on the reflection locus but also on the normal to the line *TR* at its midpoint. High-angle-ray reflection of frequency f_1 to the

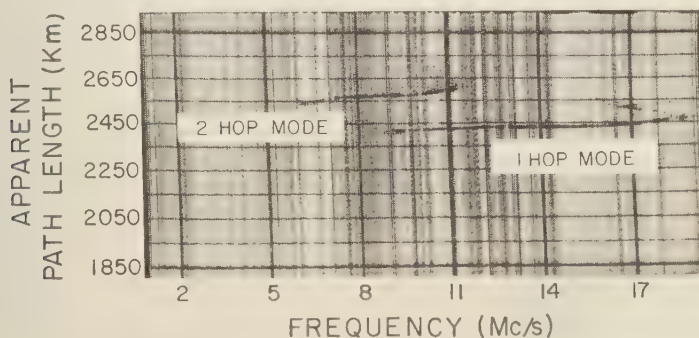


Fig. 1. Ottawa-Saskatoon, December 6, 1954, 13:41 EST.

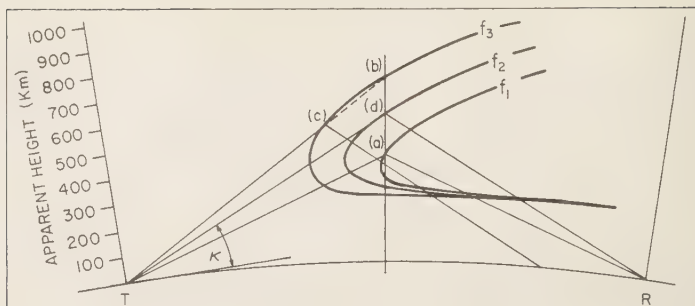


Fig. 2. The conditions determining the high-angle-ray minimum frequency.

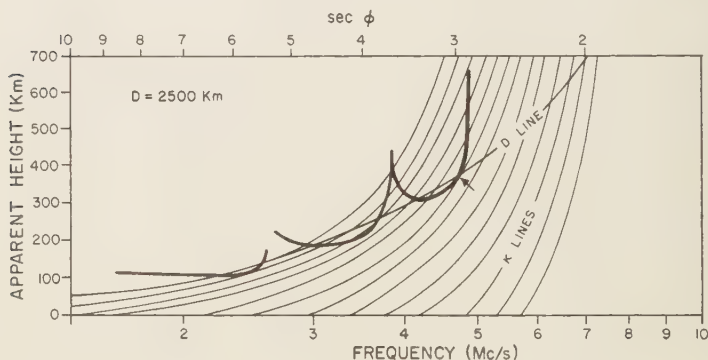


Fig. 3. The prediction of high-angle-ray minimum frequency.

receiver occurs at the point (a). High-angle-ray propagation is not possible for the frequency f_3 at (b) because of prior reflection at the point (c) which returns the wave to earth short of the receiver. The frequency f_2 is the lowest one for which high-angle-ray propagation can occur. It is the frequency for which the apparent ray path is tangent to the reflection locus at the point of reflection. This tangency condition for the lower frequency limit may be expressed by the relation

$$dk/dh' = 0 \quad (2)$$

The lowest frequency for high-angle-ray propagation can be predicted from vertical-incidence records by superimposing on them suitably chosen curves of constant $\kappa(\sec \phi, h')$ and $D(\sec \phi, h')$ in such a way as to represent the conditions described above; curves of these types are available in Figure 6.7 of the *National Bureau of Standards Circular 462* [1948]. The transmission slider $D(\sec \phi, h')$ for the given

separation of transmitter and receiver and the curves of constant κ that intersect this D curve over a reasonable range of heights are relevant to the calculation. The transmission slider and ionogram are so positioned that a κ line is tangent to the $h'(f_s)$ curve at a point that lies on the D curve. This technique is illustrated in Figure 3 for the F_2 layer. The lowest frequency propagated by this mode is computed from the coordinates of this point of tangency using the secant law. The combination of the D line and the κ lines constitutes a slider for the prediction from vertical-incidence records of the lowest frequency for the high-angle mode of propagation.

It will be seen that a single curve based on the same restrictions can take the place of the slider described above. To satisfy the restrictions, this curve must have a slope, $d(\ln f_s)/dh'$ such that dk/dh' is zero at all heights. When this curve is placed tangent to the curve of $h'(f_s)$, the point of tangency will yield the one point of the $h'(f_s)$ curve for which dk/dh' is zero. In

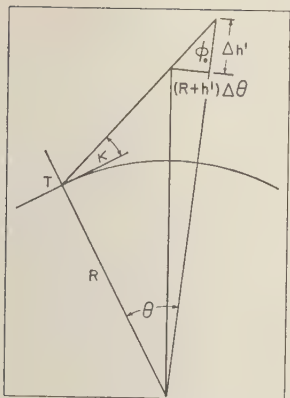


Fig. 4. The geometry showing the value of $(d\theta/dh')_x$.

order to obtain $d(\ln f_s)/dh'$ as a function of κ/dh' , it is convenient to write the secant law in the form

$$f_s = f \sin(\kappa + \theta - \xi) \quad (3)$$

where

$$\kappa + \theta + \phi_0 = \pi/2 \quad (4)$$

and

$$\xi = \phi - \phi_0 \quad (5)$$

The angles κ , θ , and ϕ_0 are shown in Figure 4. From equation 3,

$$\frac{df_s}{dh'} = f \cos(\kappa + \theta - \xi) \left[\frac{d\kappa}{dh'} + \frac{d\theta}{dh'} - \frac{d\xi}{dh'} \right] \quad (6)$$

For the condition that $d\kappa/dh'$ is zero,

$$\frac{df_s}{dh'} = f \cos(\kappa + \theta - \xi) \left[\left(\frac{d\theta}{dh'} \right)_x - \left(\frac{d\xi}{dh'} \right)_x \right] \quad (7)$$

A value for $(d\theta/dh')_x$ is obtained by reference to Figure 4:

$$(d\theta/dh')_x = (\tan \phi_0)/(h' + R) \quad (8)$$

The substitution of this value in equation 7 and the combination of the result with equations 3, 4, and 5 produce

$$\frac{d}{dh'} (\ln f_s) = \tan \phi \left[\frac{\tan \phi_0}{h' + R} - \left(\frac{d\xi}{dh'} \right)_x \right] \quad (9)$$

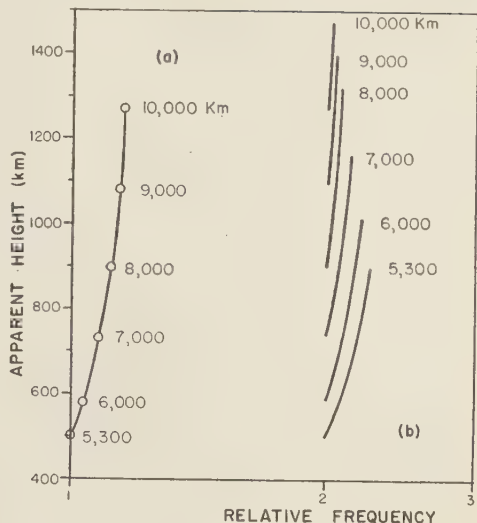


Fig. 5. High-angle-ray tangency sliders for (a) maximum distance, (b) minimum frequency.

This equation therefore gives the slope of a curve such that $d\kappa/dh'$ is zero at all heights. The required sliders could be obtained easily from equation 9 by numerical integration were it not for a difficulty in the evaluation of $(d\xi/dh')_x$.

As is indicated in equation 5, ξ is the difference between the corrected and the geometrical values of the angle of incidence. It clearly varies with height. However, existing sets of transmission sliders are far from being accurate enough to provide a reliable indication of the magnitude of $(d\xi/dh')(h', D)$. Since ξ is itself a first-order correction, the assumption that $d\xi/dh'$ is zero is reasonably accurate for the shorter transmission distances (up to 2500 km, say, for F-layer propagation). It will be necessary to revise the transmission curves for the longer distances in order to obtain accurate predictions.

In integrating equation 9 to obtain a single slider to replace the combination of D and κ lines, the quantity $d\xi/dh'$ has been taken as zero for both short and long distances. The consequences will be apparent in the comparison of the prediction with the experimental results. This slider locates the point on the $h'(f_s)$ curve from the coordinates of which the lower frequency limit may be computed by means of the secant law. The resulting sliders for a series of trans-

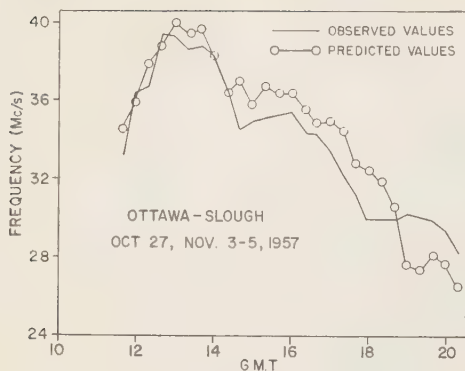


Fig. 6. The minimum frequency for two-hop high-angle-ray F_2 propagation, Ottawa-Slough, November 1957.

mitter—receiver separations are presented in Figure 5b.

The minimum frequency for two-hop high-angle-ray propagation can be predicted using a combination of D lines and κ lines in a way similar to that described for propagation by a single hop. The two vertical incidence ionograms

for the locations on the circuit where the ionospheric reflections occur are plotted on the same sheet of graph paper. The slider consisting of the D lines and κ lines is positioned so that a κ line is tangent to one $h'(f_o)$ curve while the points of intersection of this κ line on the two $h'(f_o)$ curves are traversed by two D lines for which the distances add to the transmitter-receiver separation. Under these conditions, there are usually two points of intersection of the κ line with one of the $h'(f_o)$ curves. The point of intersection that occurs at the greater height should not be used. The point of tangency of the κ line and the $h'(f_o)$ curve gives a value of f_o and of $\sec \phi$ which are multiplied to obtain the required minimum frequency.

The lower frequency limit for two-hop high-angle-ray propagation between Ottawa and Slough was predicted for October 27 and November 3-5, 1957. Vertical incidence records from the St. John's ionospheric station were used. Predicted and observed values are presented as a function of the time of day in Figure 6.

Because the density of the ionization in the

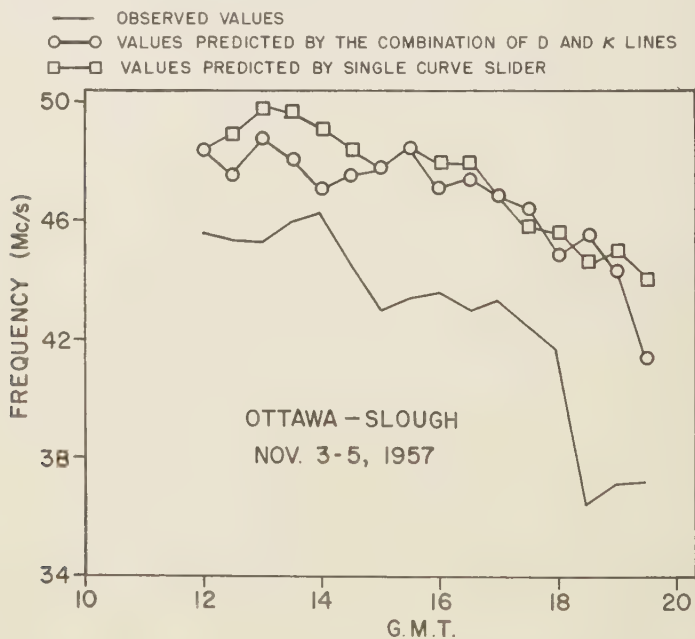


Fig. 7. The minimum frequency for one-hop high-angle-ray F_2 propagation, Ottawa-Slough, November 1957.

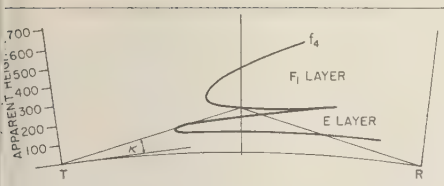


Fig. 8. The conditions determining the low-angle-ray minimum frequency.

obscurer layer is dependent on latitude, and since the data for the appropriate latitudes were not always available, a latitude correction was made, where relevant, on all the predictions shown in this paper. For one-hop high-angle-ray limits, this correction consists of the percentage difference between the $F_2 - 4000$ MUF for the latitude of the St. John's ionospheric station and that for the latitude of the midpoint of the circuit. For the two-hop high-angle-ray limits represented in Figure 6, the correction is the percentage difference between the critical frequency or the latitude of the St. John's station and that or the latitude of the reflection location where the ionosphere has the lower critical frequency. The ionogram representing this location is the one to which the κ line of the slider becomes tangent.

Predicted and observed values for one-hop propagation between Ottawa and Slough for the interval November 3 to November 5, 1957, are shown in Figure 7. Predictions were made using the two kinds of slider. For these one-hop predictions, the correction factor has been averaged over the range of latitudes that the wave travels through the ionosphere. There is little difference in the values predicted by the two sliders, but they differ appreciably from the observed values. This supports the assertion made previously, that the transmission sliders for long distances need improvement. Accurate sliders are being computed for the full range of distances for electron density distributions encountered most frequently.

The low-angle ray. The conditions determining the lowest frequency propagated by the low-angle ray may also be established with the aid of the reflection locus concept. A reflection locus constructed for E -layer and F_1 -layer heights is given in Figure 8. Low-angle F_1 -layer propagation to the receiver R occurs at the lower of the two points of intersection of the

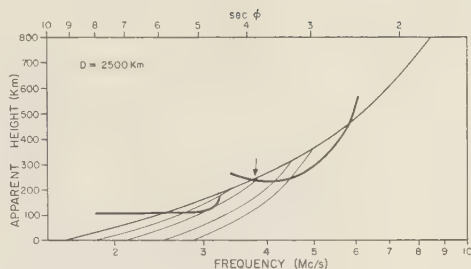


Fig. 9. The prediction of low-angle-ray minimum frequency.

reflection locus with the normal to TR at its midpoint. If the apparent ray path intersects the E -region part of the reflection locus, the ray is reflected at some E -region height with the result that the low-angle F_1 -layer reflection is obscured. Frequency f_4 is the lowest frequency propagated by the F_1 low-angle ray on the circuit TR . It is the frequency for which the apparent ray path to the F_1 -layer reflection point is tangent to the E -region part of the reflection locus.

These conditions are expressed by equations

$$f_{vE} \sec \phi_E(\kappa) = f_{vF_1} \sec \phi_{F_1}(\kappa) \quad (10)$$

$$(d\kappa/dh')(\kappa) = 0 \quad (11)$$

subject to the requirement that they be satisfied simultaneously for a single κ value. Equation 10 is true for frequencies that can be reflected from the E layer at some point along the apparent path that the ray otherwise would take for reflection to the receiver from the F_1 layer. Equation 11 is the tangency condition that must hold at some point in the obscuring layer at the elevation angle for which the wave just penetrates it if the number density of electrons in the obscuring layer has a maximum value.

The lowest frequency propagated by the F_1 low-angle ray for a given circuit can be predicted from an ionogram using the slider consisting of the relevant D line and constant κ lines which was described previously. The slider is superimposed on the ionogram so that one of the constant κ lines is tangent to the E -layer $h'(f_e)$ curve and also intersects the F_1 -layer $h'(f_e)$ curve and the D curve at their point of intersection. Figure 9 illustrates the use of the slider for one-hop propagation on a circuit of length 2500 km. This slider technique involves the assumption

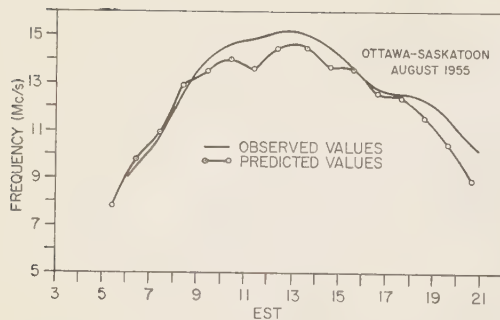


Fig. 10. The minimum frequency for one-hop low-angle-ray F_1 propagation, Ottawa-Saskatoon, August 1955.

that horizontal gradients of electron density in the ionosphere can be ignored.

It is easily seen that the same arguments are applicable to the lower frequency limit for F_2 low-angle-ray propagation. Ordinarily it is the F_1 layer, when it exists, that obscures the F_2 layer. Under certain circumstances, however, the E layer obscures the F_2 layer at a higher frequency than that at which the F_1 layer would have obscured it.

For a propagation mode involving n hops, there are $2n$ passages through the obscuring layer and therefore $2n$ possible obscuring points. The obscuring point that gives the highest predicted frequency is the one that is used in this method; it is the one nearest the subsolar point when there is no difference in latitude between the points. The vertical-incidence data must be selected for the appropriate local time and latitude.

The lowest frequencies for the one-hop low-angle F_1 mode of propagation between Ottawa and Saskatoon were predicted from Winnipeg $h'(f_o)$ records for August 1955. They are compared with observed values in Figure 10. Figure 11 shows observed and predicted lower frequency limits for the three-hop low-angle-ray F_2 mode between Ottawa and Slough for November 1957. Data from St. John's, Newfoundland, were used for this prediction.

An accurate single slider for the prediction of the lower frequency limit of the low-angle ray is not possible because its form would be a function of the apparent heights of reflection in both the obscuring and the reflecting layers. Equation 10 when rearranged takes the form

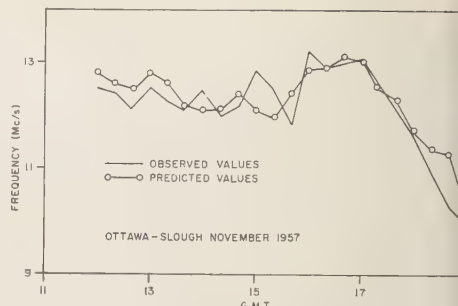


Fig. 11. The minimum frequency for three-hop low-angle-ray F_2 propagation, Ottawa-Slough, November 1957.

$$\frac{f_{zF}}{f_{zE}} = \frac{\sec \phi_E(h', \kappa)}{\sec \phi_F(h', \kappa)} \quad (12)$$

The right-hand side of (12) may be considered to be a function of two independent variables h_E' and h_F' . If a constant value for h_E' is assumed, the ratio f_{zF}/f_{zE} can be found and plotted as a function of h_F' . When this function is plotted on a logarithmic scale, the resultant curve constitutes a slider which multiplies the ratio by f_{zE} . If this single curve slider is superimposed on an $h'(f_o)$ record with scales compatible, equations 10 and 11 are satisfied by the coordinates of the intersection points when the slider is tangent to the $h'(f_o)$ curve for the E layer. The calculation of the lower frequency limit is self-consistent only when the assumed E -layer height is the height at the point of tangency. A slider based on an average E -layer height will give approximate results for all ionograms. A similar slider would result from plotting the ratio as a function of E -layer height and assuming a constant F -layer height. The latter slider will not be expected to yield as good results.

The use of the combination of D and κ lines from *Bulletin 462* does not, like the use of the single sliders shown, involve the assumption that $(d\xi/dh')_x$ is zero. Since the value of $(d\xi/dh')_x$ assumed in the use of the D and κ lines can be established for any specific set of circumstances, it is of interest to compare the results of these two methods with the observed values to give an estimate of the value of $(d\xi/dh')_x$ required to give an accurate prediction.

For most circuits lower frequencies can be propagated by low-angle than by high-angle ray. For sufficiently great transmission distance or each hop, the earth intercepts a certain amount of

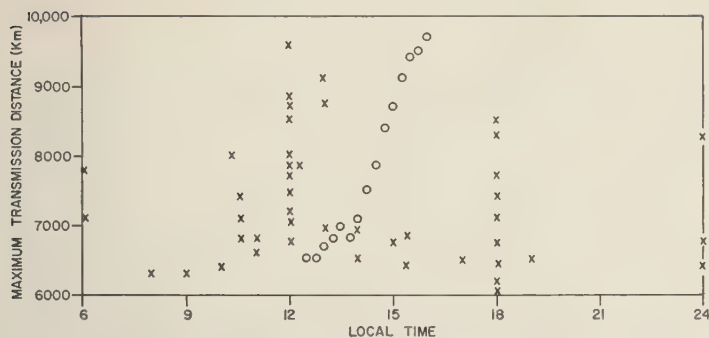


Fig. 12. Maximum one-hop transmission distances for transequatorial propagation. Stations represented: Huancayo, Talara, Ahmedabad, Ibadan, Singapore, Kodiakanal, Nairobi, Leopoldville.

low-angle-ray propagation; consequently lower frequencies can be propagated by the high-angle than by the low-angle ray. The mode that produces the lower minimum frequency can be determined readily by means of the sliders which have been presented.

The maximum one-hop transmission distance. Long-distance propagation by one hop occurs when the F -layer heights are sufficiently great. The distance of transmission for an ionospheric layer approaches a limit as the ray from the transmitter approaches tangency to the earth; the maximum transmission distance is greater for greater heights of reflection. For demonstrating the possibility of transmission distances greater than 6000 km in the equatorial regions, a slider, Figure 5(a), was constructed which yields the maximum transmission distance for the one-hop mode. It was obtained by the evaluation of the integral

$$\ln f_o = \int_0^{h'} \left(\frac{\tan \phi \tan \phi_o}{h' + R} \right)_{\epsilon=0} dh' \quad (13)$$

by numerical methods. The maximum distance was found by superimposing this slider on a vertical-incidence ionogram and obtaining the point of tangency; the distance is found by interpolating between the values marked on the slider. Equatorial-region ionograms were available from the *IGY Atlas of Ionograms* [1957]. Figure 12 presents the results obtained from records taken during 1956 and 1957 at stations within 14° of the geomagnetic equator. (Circled points were obtained from a sequence of Huancayo records that showed the development of a lunar stratification.)

The transatlantic one-hop results in Figure 7 show that the prediction made using the single slider is conservative in that a larger range of frequencies was observed than was predicted. It is clear, then, that the maximum distance slider, which was derived from the same assumptions, will also give conservative results. The transmission distances of Figure 12 should therefore be appreciably greater.

Conclusions. Methods for the computation of the lower limiting frequencies for F -layer propagation have been described for both the high- and the low-angle rays. A prediction system is now possible that yields the range of usable high frequencies, in which the prediction of the lowest usable frequency and gaps in the spectrum are treated with the same accuracy as the prediction of the maximum usable frequency.

Acknowledgments. Most of the vertical-incidence ionograms were recorded by the Department of Transport for the Defence Research Telecommunications Establishment. We are also indebted to the U. S. Signal Corps for copies of the St. John's ionograms.

REFERENCES

- I. G. Y. *Atlas of Ionograms, 1957-1958*, compiled by the Central Radio Propagation Laboratory, National Bureau of Standards, Boulder, Colorado, June 1957.
- Lejay, P., and D. Lepechinsky, *Nature*, **165**, 306, 1950.
- National Bureau of Standards, Ionospheric radio propagation, Central Radio Propagation Laboratory, *Natl. Bur. Standards Circ. 462*, **74**, June 25, 1948.
- Petersen, A. M., *J. Geophys. Research*, **56**, 221, 1951.

(Manuscript received September 17, 1959.)

A Model of the F Region Above $h_{\max}F2$

J. W. WRIGHT

National Bureau of Standards
Boulder, Colorado

Abstract. A simple Chapman model (scale height 100 km) of the ionosphere above the peak of the F region is found to be in good agreement with the few observed profiles of this region. The ratio of the above-peak electron content implied by the model to below-peak electron content is found to vary from about 2.8 to 4.0, also in good agreement with observations. The model is used in conjunction with electron density profiles observed from the ground to construct meridional cross sections along the 75°W geographic meridian.

Introduction. Conventional, ground-based ionospheric soundings are capable of providing directly the electron density profile of only that part of the ionosphere below the F_2 layer peak ($h_{\max}F2$). For many applications, however, it is desirable to know something about the distribution of ionization above as well as below $h_{\max}F2$. For example, moon echo experiments or ionosphere studies using satellite radio signals often measure the total electron content by Faraday rotation methods, and comparison with ground-based soundings requires that such an extrapolation be made. Also, in obtaining average electron density profiles from many days' data involving a range of values of h_{\max} , an extrapolation is necessary if meaningful values are to be obtained near the level of average h_{\max} .

The observations providing a basis for such extrapolation are as yet very sparse. Particularly if the occasional references to the existence of a G layer above the F_2 layer are an indication, any simple extrapolation would be incorrect. Recently available data and some theoretical considerations lend support to the reasonably simple model described below.

The choice of a model. It has been shown by several workers [Yonezawa, 1955; Hirono, 1955; Dungey, 1956; Martyn, 1956] that an ionosphere layer in which the effective processes are electron-ion diffusion, electron loss by recombination-dependent attachment, and uniform or zero vertical drift rapidly assumes a shape characterized by the well known Chapman expression.

$$N = N_{\max} \exp \frac{1}{2}(1 - z - e^{-z}) \quad (1)$$

where N_{\max} is the peak electron density and $z = (h - h_{\max})/H$ is the normalized height measured from h_{\max} in units of the scale height H .

It should be noted that the processes of diffusion and electron loss by attachment are quite different from the processes assumed by Chapman in originally deriving equation 1. The expectation that this expression applies to the upper F region is therefore not based on the fulfillment of Chapman's assumptions but on the following generally accepted characteristics of this region:

(a) Electron loss follows an attachment law with a loss coefficient given by Ratcliffe and others [1956]

$$\beta = 10^{-4} \exp \left(\frac{300 - h}{50} \right) \text{sec}^{-1} \quad (2)$$

(b) The constituent gases are in complete diffusion equilibrium at altitudes exceeding approximately 300 km. Indeed, Ratcliffe and his co-workers (*loc. cit.*) suggest that the level h_{\max} may be determined by the diffusion process itself.

(c) Electron production in the F region reaches its peak near 180 km, and therefore several scale heights below the level $h_{\max}F_2$ [Ratcliffe and others, 1956]. If electron production follows the Chapman law

$$q = q_0 \exp \left(1 - \frac{h - h_0}{H} - \sec \chi e^{-\frac{h - h_0}{H}} \right) \quad (3)$$

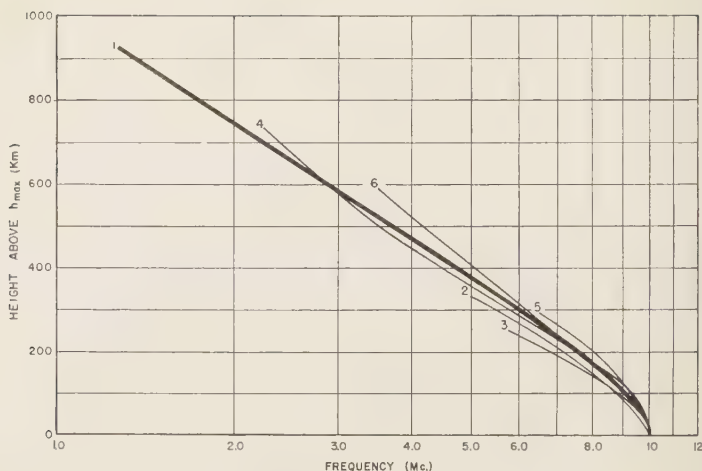


Fig. 1. Comparison of observed profiles with Chapman model ($H = 100$ km) above h_{\max} . All curves normalized to $f_{\max} = 10$ Mc/s. See text for data on curves.

where q_0 is the rate of production at the height h_0 for solar zenith angle $\chi = 0$, then during the daylight hours electron production at altitudes above $h_{\max}F2$ is relatively small, and nearly independent of χ . Therefore, the neglect of electron production in the diffusion theory leading to equation 1 does not invalidate the use of equation 1 in the daytime.

(d) Vertical drift does occur, but is probably independent of height in the F region [Maeda, 1955].

The parameters that determine the distribution given by (1) are N_{\max} and h_{\max} , both obtainable from the ground-based soundings, and H , the scale height. From the diffusion theory, this should be the scale height of the substance ionized in the F region, which is thought to be atomic oxygen. From theoretical considerations, Nicolet [1954] gives $H = 99$ km at 400 km as a probable value for atomic oxygen. We have thus chosen $H = 100$ km for the present model.

Comparison with observed profiles above $h_{\max}F2$. The Chapman function with $H = 100$ km is shown in Figure 1. For comparison, five other 'topside' profiles observed in various ways are also shown. All have been adjusted to the same h_{\max} and normalized to a value of N_{\max} corresponding to $f_oF2 = 10$ Mc/s.

Curve 1 is the Chapman model, with $H = 100$ km.

Curves 2 and 3 are measurements of the electron density profile over Havana, Illinois, by the 'scatter sounding' method; 1135 and 194 CST, February 27, 1959 (Bowles, private communication).

Curve 4 is the 'probable profile' deduced by the Naval Research Laboratory from satellite and rocket data of the USSR. These data represent an estimated average ionosphere for summer noon in middle latitudes (Seddon, private communication).

Curve 5 is the smoothed electron density profile deduced by the Space Technology Laboratories from a rocket firing over Cape Canaveral, August 2, 1958, at 1718 EST [Farmer and Robinson, 1959].

Curve 6 is the profile deduced by Alpert and others [1958] from radio observations of Sputnik 1, October 5, 6, and 7, 1957.

A fair degree of agreement is seen to exist among all these observations, substantiating the choice of the Chapman model with $H = 100$ km as a reasonable extrapolation for the F -region ionization above h_{\max} . The five observations which are well distributed seasonally, do not suggest a seasonal variation.

Comparison with integrated electron content measurements. Measurements of the (Faraday) rotation of the plane of polarization of satellite radio signals or radar moon echoes can

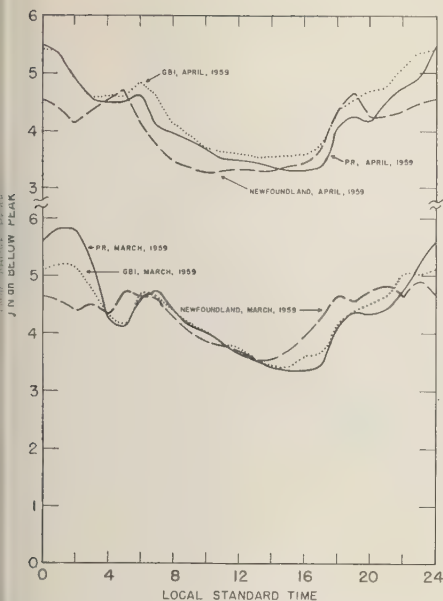


Fig. 2. The effective ratio of above-peak to below-peak total electron content implied by application of the Chapman model to observed electron density profiles.

provide estimates of the total electron content along the path of the signal; more often a comparison with ground-based soundings gives the ratio of the integrated electron content above the maximum to that below [Bauer and Daniels, 1958]. To provide a means for comparison with such observations, it is of interest to consider several properties of the Chapman model with regard to its integrated electron content.

The integrated electron content from the ground to a level h of the ionized layer described by (1) is given by

$$\int_{-\infty}^h N dh = \sqrt{2\pi e} HN_{\max} [1 - P(x)] \quad (4)$$

where $P(x)$ is the probability integral

$$P(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du$$

and x is related to the height h by

$$x = \frac{1}{\sqrt{2}} \exp \left(-\frac{1}{2} \frac{h - h_{\max}}{H} \right)$$

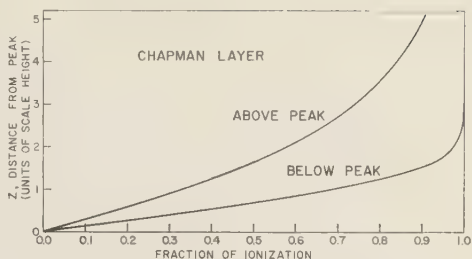


Fig. 3. Fraction of electron content between peak and level z .

Some useful special cases from this expression are:

- Total integrated content
 $N_t = \sqrt{2\pi e} HN_{\max} = (4.133) HN_{\max}$.
- Integrated content below h_{\max} ,
 $N_b = (1.312) HN_{\max}$.
- Integrated content above h_{\max} ,
 $N_a = (2.821) HN_{\max}$.
- $N_a/N_b = R = 2.15$.

The use of this model permits us to calculate an effective ratio R_e of the electron content above h_{\max} based upon the model to the observed content below h_{\max} . R_e is usually larger than the value 2.15 appropriate to the complete Chapman distribution. The effective ratio is given by

$$R_e = (N_a/Sh_{\max}) = 2.821 (HN_{\max}/Sh_{\max})$$

where Sh_{\max} is the observed electron content below h_{\max} . The quantity R_e may thus be obtained from observed electron density profiles: N_a involves only the assumed scale height and the peak density of the layer, and Sh_{\max} is obtained by integration of observed profiles. Quiet-day mean profile data for the months of March and April 1959, from Newfoundland, Grand Bahama Island, and Puerto Rico, were used to compute the values of R_e shown in Figure 2. These curves are fairly consistent in implying an effective ratio of about 3 in the daytime and (with less consistency) of about 4 or 5 at night.

For comparison, Table 1 shows estimates of the ratio from moon echo and satellite observations.

The effective ratio of electron content as defined above and shown in Figure 2 compares quite well with the observational values of Table 1. Furthermore, N_a/Sh_{\max} shows a diurnal

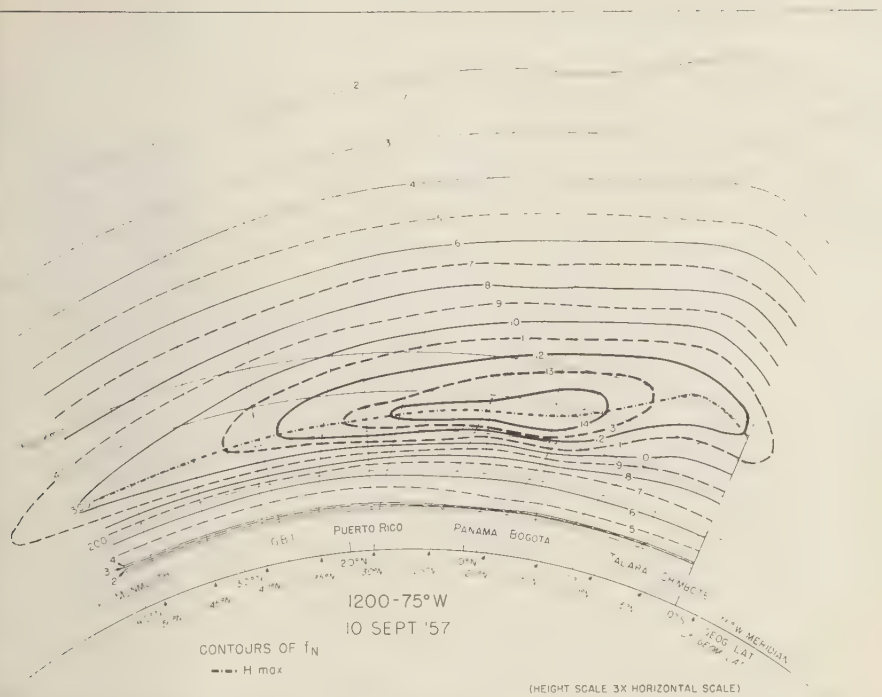


Fig. 5. Ionospheric section along the 75°W geographic meridian, showing results of extrapolation above h_{\max} .

tions located along the 75°W geographic meridian. Results of this study will be described elsewhere, but it is of interest here to apply the 'topside' extrapolation described above to the electron density profiles obtained in this study. Figures 4, 5, and 6 illustrate the resulting cross section of the ionosphere at three different longitudes of a geomagnetically quiet equinoctial day. The contours are of plasma frequency; they therefore show the level of reflection of the corresponding radio frequency at vertical incidence. Both geographic and geomagnetic coordinates are shown, and the lines of the magnetic field are indicated for each 5° of magnetic latitude. Note that the vertical scale is expanded three times.

As would be expected, most of the gross irregularity evident at the h_{\max} level has disappeared at heights of 1000 km or more, and it might seem likely that the remaining irregularities are suspect. However, the conclusions of

Hirano [1955], Dungey [1956], and Martyn [1956] that a Chapman distribution also develops under the processes of diffusion and decay would indicate that the night-time profiles of Figures 4 and 6 are probably approximately correct, except possibly just at the equator where the vertical diffusion process is inhibited by the horizontal magnetic field.

Martyn [1959] has also conjectured that the regions of intense ionization about 17° from the equator (represented in Figures 4 to 6 by the ionosphere over Bogotá) are probably caused by nearly horizontal diffusion along the magnetic field from the high, thick daytime F region over the geomagnetic equator. It is interesting to note that at night the present model gives (perhaps fortuitously) above-peak ionization contours that are consistent with this mechanism. In Figures 4 and 6, north of Bogotá, the contours are nearly parallel to the geomagnetic field; toward the equator they tend to be nearly

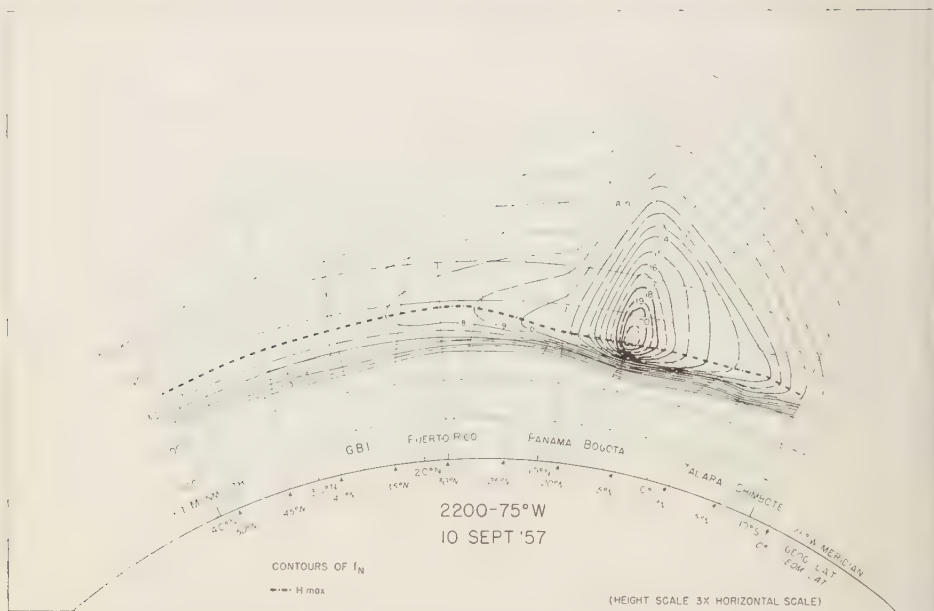


Fig. 6. Ionospheric section along the 75°W geographic meridian, showing results of extrapolation above h_{\max} .

perpendicular to it, suggesting that the process primarily responsible for the extra ionization at night over Bogotá is indeed transport from regions nearer the equator.

Acknowledgments. The guidance and advice of Mr. T. N. Gautier are gratefully acknowledged, as are the benefits of many helpful discussions with Dr. T. E. Van Zandt, both of this laboratory.

REFERENCES

- Alpert, Y. L., E. F. Chudsenro, and B. S. Shapiro, Results of research on the outer regions of the ionosphere, from observations of the radio signals of Sputnik 1, translated by E. R. Hope, Defence Research Board, Ottawa, Canada, October 1958, from *Preliminary results of scientific researches on the first Soviet artificial satellites and rockets, No. 1*, published by the Academy of Science of the USSR, Moscow, pp. 40-108, 1958.
- Bauer, S. J., and F. B. Daniels, Ionospheric parameters deduced from the Faraday rotation of lunar radio reflection, *J. Geophys. Research*, 439, 1958.
- Dungey, J. W., The effect of ambipolar diffusion on the nighttime F layer, *J. Atmospheric and Terrest. Phys.*, 9, 90-102, 1956.
- Evans, J. V., The measurement of the electron content of the ionosphere by the lunar radio echo method, *Proc. Phys. Soc. London, B*, 953, 1956.
- Evans, J. V., The electron content of the ionosphere, *J. Atmospheric and Terrest. Phys.*, 259, 1957.
- Farmer, D. J., and W. A. Robinson, Density distribution of free electrons in the upper atmosphere, presented at spring meeting of UR Washington, D. C., 1959.
- Hirano, M., Effect of gravity and ionization pressure gradient on the vertical drift in the region, *Rept. Ionosphere research Japan*, 95-104, 1955.
- Maeda, K., Theoretical study on the geomagnetic distortion in the F2 layer, *Rep. Ionosp. F Japan*, 9, 71-85, 1955.
- Martyn, D. F., Processes controlling ionization distribution in the F2 region of the ionosphere, *Australian J. Phys.*, 9, 161-165, 1956.

Artyn, D. F., The normal *F* region of the ionosphere, *Proc. IRE*, 47, 147, 1959.

Colet, M., Dynamic effects in the high atmosphere, in *The Earth as a Planet*, edited by G. P. Kuiper, University of Chicago Press, pp. 644-12, 1954.

Cliff, J. R., E. R. Schmerling, G. S. G. K. Getty, and J. O. Thomas. The rates of production and loss of electrons in the *F* region of the

ionosphere, *Phil. Trans. Roy. Soc. London, Ser. A*, 248, 621-642, 1956.

Yonezawa, T., On the influence of electron-ion diffusion on the electron density and height of the nocturnal *F2* layer, *J. Radio Research Labs.*, 2, 125-136, 1955.

(Manuscript received July 28, 1959; revised October 16, 1959.)

An Analysis of Time Variations in Tropospheric Refractive Index and Apparent Radio Path Length¹

M. C. THOMPSON, JR., H. B. JANES, AND A. W. KIRKPATRICK

National Bureau of Standards, Boulder, Colorado

Abstract. The National Bureau of Standards has been conducting a series of measurements for a study of the characteristics of the turbulent lower atmosphere and its effect on the accuracy of radio direction-finding, guidance, and geodetic measurement systems. The results of three experiments are presented, one consisting of recordings of refractive index and apparent path length variations at 9400 Mc/s over a 15.5-mile path on Maui, Hawaii and the other two consisting of similar measurements made over a 9.5-mile path near Boulder, Colorado. The correlation of refractive index and apparent path length fluctuations is discussed as well as the power (variance) density spectra of both variables.

Introduction. In recent years various radio systems have been developed for the accurate measurement of distance and/or velocity. Their applications range from missile tracking and guidance to geodetic measurements. One feature which these systems have in common is the requirement that an estimate be made of the refractive index of the atmosphere along the propagation path. Since this fluctuates as a function of both time and space as a result of turbulence and other mechanisms, the atmosphere, in effect, limits the ultimate accuracy of such measurements.

The National Bureau of Standards, under the sponsorship of the Air Force Ballistic Missile Division, has made a series of measurements for a study of the nature and extent of this atmosphere-induced limitation. The data obtained are of two general types: the time variations in refractivity² observed at one or more points along a radio propagation path, and the corresponding variations in the phase-of-arrival of a microwave signal transmitted over the path. The measurement equipment and techniques have been described previously [Herbstreit and

Thompson, 1955; Thompson and Vetter, 1958]. The purpose of this paper is to present some statistics obtained from these experiments, particularly the power density spectrum of phase and refractivity variations taken under widely differing climatic conditions.

General description of measurements. Two series of measurements will be described in this paper. The first, made during the fall of 1956, were over a path 15.5 miles long extending from the 10,000-foot summit of Haleakala on the island of Maui, Hawaii, to a point about 100 feet above sea level near the base of the mountain. Plan and profile views of the path are shown in Figures 1 and 2. Continuous recordings were made of the phase-of-arrival of a 9400 Mc/s signal sent over the path. Vertically polarized, 18-inch parabolic antennas were used at both terminals. At the lower terminal of the Maui path an earthen wall was erected in the antenna foreground in order to block ground reflections. At the summit terminal the ground elevation drops sharply in front of the antenna. Long-term variations in refractivity were computed from temperature, pressure, and relative humidity recordings made at both path terminals and at the three intermediate points shown in Figure 1. At the summit site a microwave refractometer was used to obtain recordings of the short-term variations in refractivity.

The second series of measurements was obtained during the fall of 1958 over a 9.5-mile path near Boulder, Colorado. Plan and profile views of this path are shown in Figures 3 and 4.

¹ This is a somewhat revised and expanded version of a paper entitled "An Analysis of Observed Atmosphere-Induced Phase Variations in Line-of-Sight Microwave Propagation" presented at the Joint IRE-URSI meeting in Washington, D. C., May 4-7, 1959.

² If n is the refractive index, the refractivity, N , defined by

$$N = (n-1) \times 10^6$$



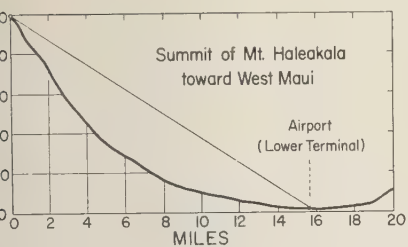
Fig. 1. Location of propagation path used in Maui experiment.

The principal difference between this path and the Maui path, aside from a difference in length and an obvious difference in climate, is that the Maui path is tilted, the angle of elevation being about 7° , compared with about 11° for the Colorado path. At both terminals of the Green Mountain Mesa-Table Mesa path, the ground elevation drops sharply in front of the antennas. In Figure 3, it appears possible that ground reflections might occur in the vicinity of Roberts Lake, a pond situated on a rather prominent mesa. However, any such reflection would probably contribute only a small part to the received signal because the lake lies well below the first Fresnel zone. Two sets of measurements were made on this path, one in September using 18-inch parabolic antennas and lasting about 40 hours, the other in October using 4-foot parabolic antennas and lasting 5 days. Horizontal polarization was used for the most part during

both runs, but the effect of polarization was investigated on this path and no change in the statistics of the phase variations could be detected when the antenna polarization was shifted from horizontal to vertical in repeated trials.

The data obtained during the September run consisted of a continuous chart recording of long-term variations in phase-of-arrival, supplemented periodically by high-speed chart recording of the short-term phase variations obtained by the use of a band-pass filter. The refractivity data were computed from meteorological data recorded at the path terminals. The analysis of these data has previously been discussed in detail. [Thompson and Janes, 1959].

During the October run the phase measurements were similar to those made in September except for an adjustment in the dynamic range of the continuous record to give better resolution. In addition to the standard meteorologic



Terrain profile of Maui propagation path.

measurements made at both terminals, a refractometer was used at Table Mesa to record refractivity variations with both low-pass and band-pass recording circuits.

Correlation of path length and refractivity variations. The correlation between apparent path length (phase-of-arrival) variations and the refractivity observed at one point on the path will, of course, depend upon the extent to which the latter is correlated with the refractiv-

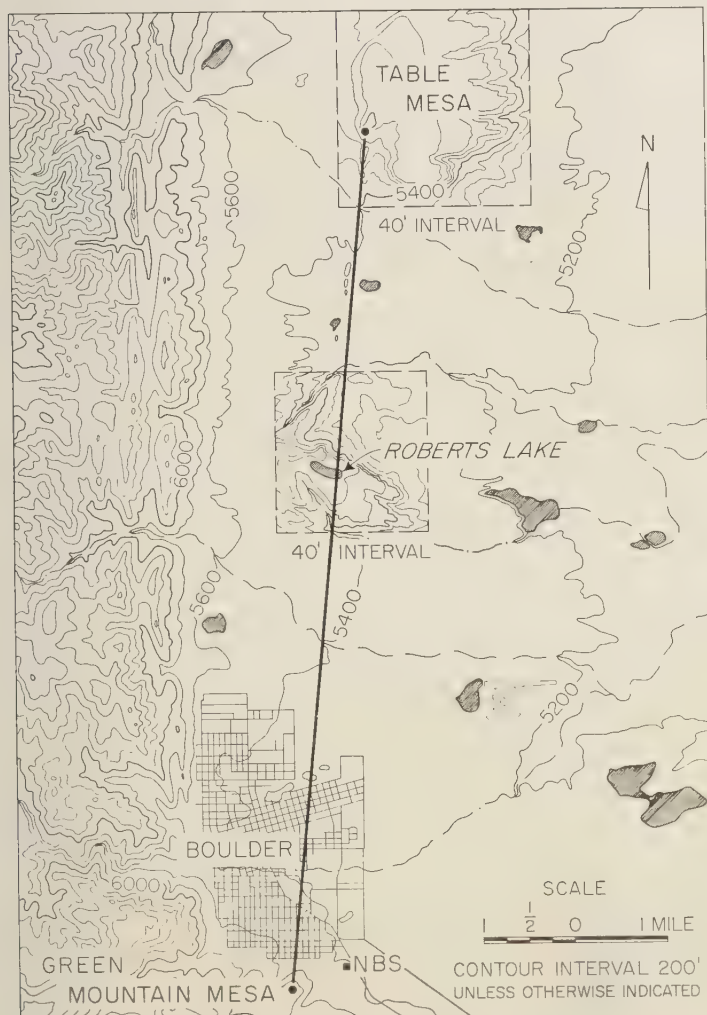


Fig. 3. Green Mountain Mesa-Table Mesa path.

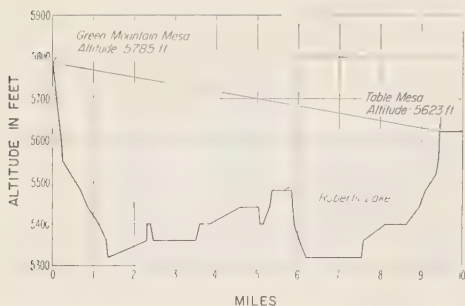


Fig. 4. Terrain profile of Green Mountain Mesa-Table Mesa path.

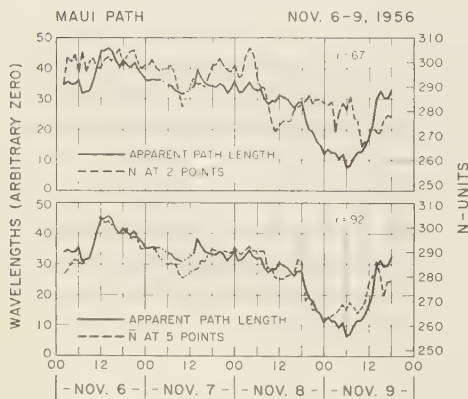


Fig. 5. Long-term changes in electrical path length and refractivity, Maui path (Nov. 6-9, 1956).

ity at other points on the path. For example, the relatively small, short-term fluctuations observed with a refractometer over periods of seconds or minutes generally reflect atmospheric conditions only in the immediate vicinity of the instrument and hence are not well correlated with the corresponding short-term path length variations. However, the larger, long-term changes in these two variables, observed over periods of several hours, are better correlated. This is illustrated in Figures 5, 6, and 7, which show for each of the three experiments the variations in apparent path length plotted with the average of refractivity observed at two or more points on the path. The ordinate scales on these graphs are adjusted so that a change in path length of 1 ppm is represented by the same displacement on the graph as a change of 1 ppm (1

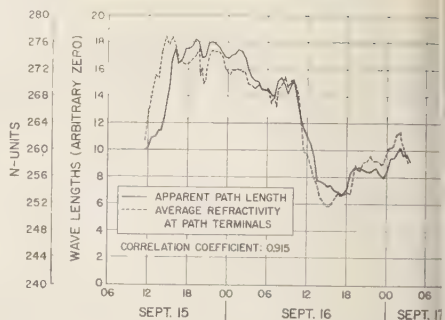


Fig. 6. Long-term changes in electrical path length and refractivity, Green Mountain Mesa-Table Mesa path (Sept. 15-17, 1958).

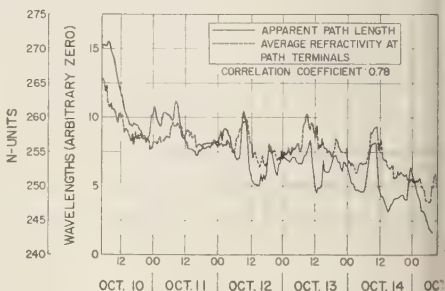
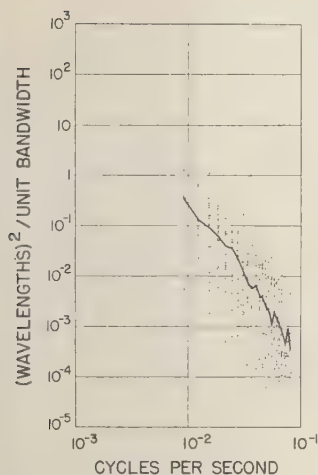


Fig. 7. Long-term changes in electrical path length and refractivity, Green Mountain Mesa-Table Mesa path (Oct. 10-15, 1958).

N -unit) in refractive index. It will be seen that there is reasonably good correlation between the long-term path length and the average of the refractivity values observed at the path terminals. As might be expected, correlation is poorer ($r = 0.67$) at Maui, where the sloping path extends through widely differing types of atmosphere which are not well represented by measurements taken at the terminals. It should be noted, however, that the correlation is much improved when the average of refractivity values taken at all five meteorological stations along the path is used.

These graphs (Figs. 5, 6, and 7) illustrate the magnitude of variations in the apparent length of a propagation path observed over a number of days. At Maui, it amounted to about 35 wavelengths (44 in.) in four days, whereas the Table Mesa path varied about 12 wavelengths (15 in.)



8. Spectrum of apparent path length variations, Maui path (Nov. 5-9, 1956).

apparent length during both the September and October runs.

Spectra of path length variations. Of the many statistics which could be used to describe the time variables, probably the most useful descriptive statistic is the power (or variance) density spectrum, which we shall refer to here as 'the spectrum.' Essentially, a plot of the spectrum shows graphically how much of the total variance of a process was contributed by variations occurring within a given frequency band. For example, suppose that a variable is proportional to the time variable. It is passed through a number of very narrow bandpass filters arranged in parallel. The spectrum could be estimated by a graph of the variance of the output of each filter divided by the filter bandwidth, plotted as a function of the center frequency.

The data discussed here were recorded on oscilloscope charts, so it was not possible to compute the spectrum by this analog method. Instead, samples of uninterrupted records were scaled numerically at equally spaced time intervals. Then the Fourier cosine transform of the autocovariance function was computed for each sample, using the general method outlined by Blackman and Tukey [1958]. This for discrete data is the equivalent of the spectrum estimate obtained by digitally filtering analog data.

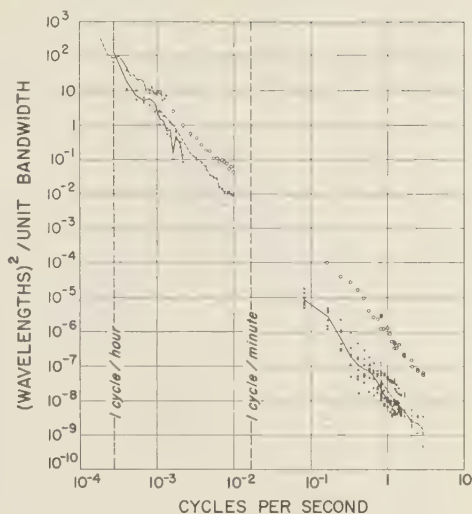


Fig. 9. Spectrum of apparent path length variations, Green Mountain Mesa-Table Mesa path (Sept. 15-17, 1958).

The result of the particular computation used here consists of a spectral estimate for each of 30 evenly spaced frequency bands whose width is inversely proportional to the data sampling interval. In the graphs which follow, these densities are plotted at the center frequency of each band. Since the pass-band of the recording circuit generally exceeded the range of frequencies covered by the spectrum computation, it was often necessary to compute the spectrum in several overlapping frequency segments, using different total sample lengths and sampling intervals. To prevent contamination of the information in a given segment by variances present at lower frequencies, the data were subjected to a trend-removal process which consists simply of computing the deviations of the variable from a running average, computing the spectra of these deviations, and correcting the result for the frequency response of this high-pass 'filter.' This, in effect, filters out variations at spectral frequencies below a cutoff frequency which is determined by the averaging time.

Figure 8 shows plots of a number of spectra computed from 15-minute samples taken from the Maui path-length data. The range of frequencies covered is comparatively small, being restricted on the high end by the resolution of

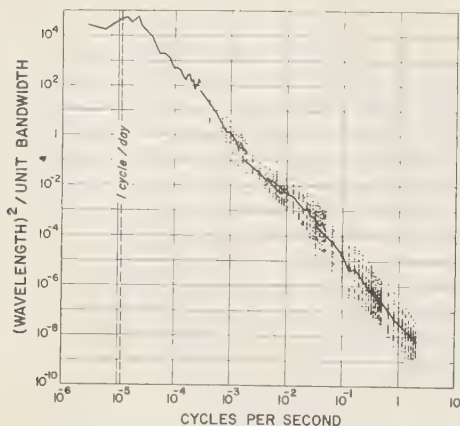


Fig. 10. Spectrum of apparent path length variations, Green Mountain Mesa-Table Mesa path (Oct. 10-15, 1958).

the recording, and at the low end by the sample length. The medians of the spectral estimates at each frequency have been joined by line segments. The two points of interest here are the slope of the spectra, (that is, the apparent frequency dependence of the spectral estimates) and their time variability at a given frequency. The slopes of the individual spectra range from -2.2 to -3.8 , with an average value of -2.9 ; that is, the spectral density is proportional to $f^{-2.9}$.

The vertical spread of the sample spectra at a given frequency is much larger than could be ascribed to random sampling fluctuations. This indicates that the variance underwent large changes during the period of recording and suggests that this is not a stationary process. The sample mean also underwent large changes, as can be seen in Figure 5. However, the variance appears to be independent of these variations in the sample mean and, in fact, to be independent of time of day, wind velocity at the path terminals, refractivity variance, or any other variable with which it has been compared.

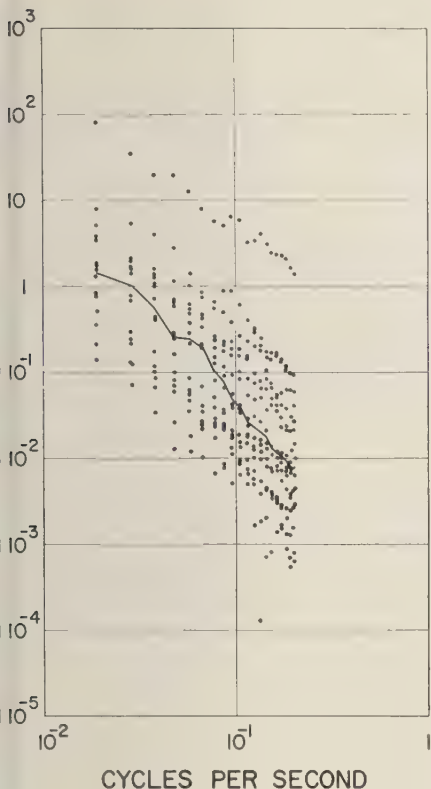
Figure 9 shows a similar plot of sample path length spectra obtained from the Table Mesa path during the 2-day run in September. The resolution and long-term continuity of the data were much improved this time, permitting extension of the range of the spectrum analysis from 5 cps to about 1 cycle per hour, except for a gap between the pass-bands of the two record-

ing systems being used. Here (and in Figs. 13 and 14) the dots and X's simply denote sets of spectral estimates computed for different sampling rates and length of record. The circles indicate estimates of the spectral values obtained during a rainstorm caused by the passage of a cold front across the path on the afternoon of September 15. This is the only instance when the power spectrum was clearly being influenced by meteorological conditions. As before, the lines join the medians of the spectral estimates (including those shown as circles) at each frequency within the somewhat overlapping frequency segments. Two interesting points here are that the average slope of the spectra is very nearly the same as at Maui, being about -2.8 , and that it remains constant over this very wide range of frequencies.

Figure 10 is a graph of sample path length spectra obtained from the Table Mesa path during a 5-day period of continuous recording in October. Because of further improvements in recording techniques, it was possible to combine more segments of the frequency scale, closing the gap in the middle of the spectrum and extending the spectrum beyond the 'hump' at 1 cycle per day. Again, there is little variation in slope over the entire range of frequencies, except at the diurnal hump. The average slope varies from segment to segment from about -2.5 to -2.7 , but if it can be assumed that the entire spectrum is a straight line, then its slope appears to be about -2.6 .

Spectra of refractivity variations. The analysis of refractivity data obtained at both Maui and Table Mesa was similar to that of the path length data. From the low-pass chart recordings made at the Haleakala summit site, 15-minute samples which were reasonably well distributed over the recording period were selected. These spectra (Fig. 11) have the same wide variability in length that was evident in the path length spectra. In the previous figures, the lines join the medians of the spectral estimates at each frequency and are intended only to serve as a guide to the eye in picturing a 'typical' spectrum. The slopes of these spectra range from -1.6 to -3.1 with an average of -2.2 , as compared with -2.9 for the corresponding path length spectra.

At Table Mesa, during the October run, the low-pass data were supplemented by high-spe-



11. Spectrum of time variations in refractivity, Haleakala summit (Nov. 5-9, 1956).

d-pass recordings made at frequent intervals in order to measure variations at higher frequencies. Figure 12 shows some sample spectra computed from both the low-pass data (for frequencies up to about 0.05 cps) and the band-pass data (for frequencies up to about 5 cps). The latter spectra tend to divide into two distinct groups, especially in the vicinity of 0.1 cps. The spectra in the upper group represent samples taken during daylight hours and the lower ones represent nighttime samples, there is temptation to estimate from this the magnitude of the diurnal variation in spectral level (that is, in variance). However, the high-level spectra were all calculated from data recorded on one day (October 14), and hence it is not clear to what extent this separation reflects diurnal or day-to-day changes. The very high densities indicated by the circles

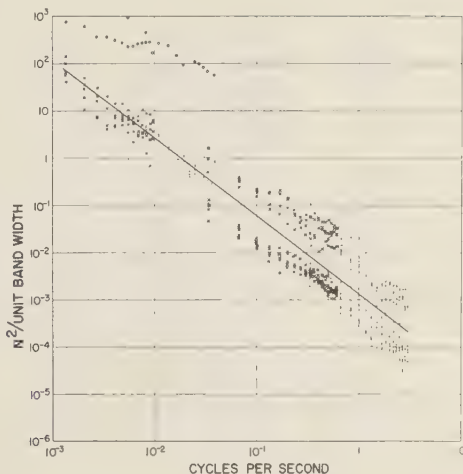


Fig. 12. Spectrum of time variations in refractivity, Table Mesa (Oct. 11-14, 1958).

are from data recorded during the day on October 12, the only period during which the wind speed exceeded 15 mph. Again, however, the evidence is too scanty to allow one to draw even tentative conclusions as to what atmospheric parameters influence the total variance of refractivity.

It should be pointed out that these spectra may be somewhat biased in the direction of lower spectral levels. This is because periods of very large fluctuations cannot be analyzed if the recording pen is repeatedly driven off scale. This is the reason that the higher frequency band-pass data could not be analyzed during the period of large fluctuations whose lower frequency components are represented by the circles in Figure 12. On the other hand, during the afternoon of October 11, the refractivity fluctuations were too small to be adequately resolved by the recording equipment and hence could not be analyzed. However, it is believed that the spectra illustrated are representative of those to be expected a large percentage of the time on this path during the autumn. The straight line drawn through the points is simply a visual estimate of a fit to all the spectra except for the one marked with circles. The slope of this line is -1.5 . The calculated slopes for the individual spectra (again excluding the one denoted by circles) range from -1.0 to -2.3 , with an average of -1.6 .

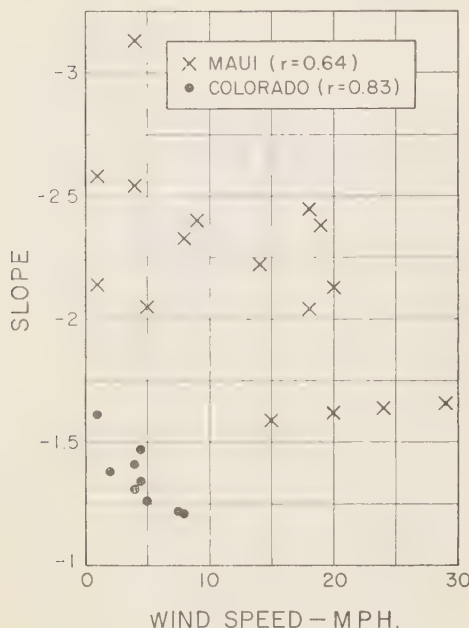


Fig. 13. Slope of refractivity spectrum versus wind speed.

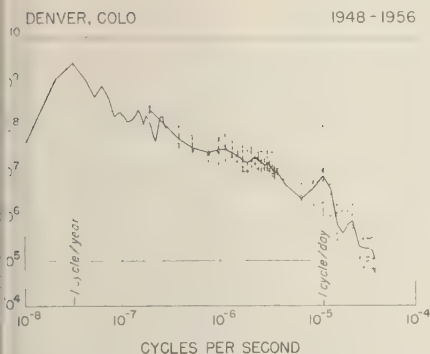
Figure 13 is a plot of the slopes of individual sample refractivity spectra plotted against wind speed. Included are the Maui spectra and those Table Mesa spectra falling in the frequency range from about 0.03 to 0.6 cps. This graph suggests that a correlation does exist, with relatively steep slopes associated with relatively light winds. This is in at least qualitative agreement with the theory of turbulence in a stably stratified atmosphere [Bolgiano, 1958].

According to this theory, the slope of the wave number (space domain) spectrum might be expected to vary from $-7/5$ in the buoyancy subrange to $-5/3$ in the inertial subrange and to even steeper slopes in the viscous dissipation subrange. The relative extent of each of these subranges along the wave number axis and the portion of this spectrum to which any given sample frequency spectrum corresponds are both at least partially dependent upon the mean wind speed (using the common approximation of converting the space spectrum to a time spectrum by multiplying the wave number values by mean wind speed to obtain the corresponding frequency

values). This expected dependence in the same direction as that indicated by Figure 13.

The spectral slopes of the Table Mesa data agree reasonably well with these theoretical values and with data obtained by Birnbaum and Bussey [1955] and by Crain [1956]. However, the Maui spectra are considerably steeper. It is not clear to what extent the difference is caused by the difference in climate (the Maui path length spectra were also somewhat steeper than their Colorado counterparts) and by the difference in altitude (the Haleakala site is 10,000 feet above sea level).

Interruptions of the Table Mesa refractivity data prohibited computation of the spectrum below about 0.001 cps. However, it appears reasonable to assume that the slope will remain constant down to a frequency of 1 cycle per day as in the case of the path length data. Some evidence to support this assumption was obtained from an analysis of U. S. Weather Bureau data gathered at Stapleton Field, Denver, about 5 miles southeast of the Table Mesa site. The data consisted of bihourly values of atmospheric pressure, temperature, and relative humidity and covered the 8 years from 1948 to 1956. The corresponding values of refractivity had already been computed for the use of the Radio Meteorology Section of the National Bureau of Standards and so were readily available. These data were analyzed for this study by computing the spectrum of refractivity variations over as wide a range of frequencies as possible. Figure 14 shows the results of this analysis. The higher frequency segment of the spectrum (denoted by X's) was computed for four periods, each 2 weeks in length. These were chosen by visual inspection, two being periods of relatively high diurnal variation during the summer months and the other two being periods of low diurnal variation during the winter months. It is felt that these represent the extremes in spectral level in this range of frequencies. The middle segment of the spectrum (denoted by dots) was made up of 16 samples, each representing day-to-day variations observed at the same time of day for 1 year. This was done for four different times of day (0200, 0800, 1400, 2000 hours) during alternate years (1949, 1951, 1953, 1955). The lowest segment is a single spectrum representing



14. Spectrum of long-term refractivity variations, Denver, Colorado (1948-1956).

3 years' data sampled at the rate of 1 point every 18.25 days.

The 'humps' in the spectrum at 1 cycle per year and 1 cycle per year are clearly evident in the graph. Judging from Figures 12 and 14, the spectrum at 1 cycle per day appears to divide the spectrum into two distinct parts. At frequencies above this point, the slope appears to be about -1.5 or -1.6 over the entire range of more than frequency decades. On the other hand, between 1 point and 1 cycle per year, the slope is very nearly -1.0 .

Comparison of apparent path length and refractivity spectra. The long-term variations in apparent path length and refractivity are reasonably well correlated, as was pointed out previously. When these variables are expressed in terms of parts per million change in total path length and parts per million change in refractive index, the variations in the two quantities are very nearly equal in amplitude. This being the case, their spectra (also in terms of parts per million) should converge at the low-frequency limit. Figure 15 summarizes the spectra from the Maui and (October) Table Mesa experiments along with the Weather Bureau data. The lines represent simply the median spectral lines taken from Figures 8, 10, 11, and 14 and the 'visual fit' taken from Figure 12. From this it appears that the path length and refractometer spectra do converge in the vicinity of 1 cycle per day and that they tie in reasonably well with the Weather Bureau refractivity spectra. At the higher frequencies, the difference of approximately 1 between the slopes of the path length and refrac-

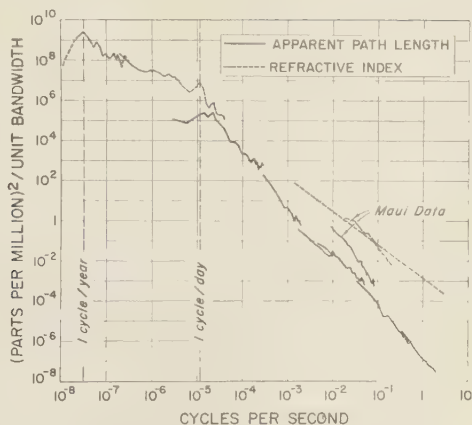


Fig. 15. Spectra of refractive index and apparent path length variations, Colorado data.

tivity spectra is in good agreement with that predicted by Norton and others (to be published).

REFERENCES

- Birnbaum, George, and H. E. Bussey, Amplitude, scale and spectrum of refractive index inhomogeneities in the first 125 meters of the atmosphere, *Proc. IRE*, **43**, 1412, 1955.
- Blackman, R. B., and J. W. Tukey, The measurement of power spectra from the point of view of communication engineering, *Bell System Tech. J.*, parts 1, 2, 37, 185, and 485, 1958.
- Bolgiano, Ralph, Jr., A meteorological interpretation of wavelength dependence in transhorizon propagation, *Cornell Univ. Research Rept., EE 385*, 1958.
- Crain, C. M., Index of refraction fluctuations and 1046 Mc phase difference measurements along an 18,000 foot elevated path, *Texas Univ. Elec. Eng. Research Lab. Rept.*, 6-15, 1956.
- Herbstreit, J. W., and M. C. Thompson, Jr., Measurements of the phase of radio waves received over transmission paths with electrical lengths varying as a result of atmospheric turbulence, *Proc. IRE*, **43**, 1391, 1955.
- Norton, K. A., J. W. Herbstreit, and M. C. Thompson, Jr., Radio studies of atmospheric turbulence, *Natl. Bur. Standards Tech. Note*, sect. 2, vol. 1 (to be published).
- Thompson, M. C., Jr., and H. B. Janes, Measurements of phase stability over a low-level tropospheric path, *J. Research Natl. Bur. Standards*, **63D**, 45, 1959.
- Thompson, M. C., Jr., and M. J. Vetter, Single path phase measuring system for three-centimeter radio waves, *Rev. Sci. Instr.*, **29**, 148-150, 1958.

(Manuscript received September 30, 1959.)

Electrical Theory of Tornadoes

BERNARD VONNEGUT

*Arthur D. Little, Inc.
Cambridge, Massachusetts*

Abstract. Modern theory and observations appear to support the very old and almost forgotten idea that tornadoes are a manifestation of thunderstorm electricity. It is suggested that there is sufficient electrical energy in an intense thunderstorm to power a tornado and that the electrification could cause extraordinarily intense winds by electrically heating air or by accelerating charged air in an electric field.

INTRODUCTION

Probably the most remarkable feature of tornadoes is the extraordinary velocity of their winds. They are commonly known to dig large furrows, twist off large trees, drive pieces of straw into wood, and lift men, cattle, and automobiles high into the air. These phenomena are unknown in other storms and testify to enormous velocities that are estimated to be from 100 m/sec to the speed of sound [Flora, 1933].

Since tornadoes are relatively small, the energy associated with them is not large by meteorological standards, despite the tremendous velocity of their winds. Some measure of the energy required to produce a tornado can be obtained if one computes the power that would be required to accelerate air continuously to a velocity of 250 m/sec over a circular area 100 meters in diameter (see Appendix I). This would require about 10^{18} ergs/sec or 10^6 kw. Though the power is somewhat in excess of the capacity of the electrical generating plants in the United States, it is small in comparison with the power available from a thunderstorm. An energy of 10^{18} ergs, which Braham [1952] has estimated for a single average thunderstorm cell, should be sufficient to provide power to run a tornado for several hours.

In a recent paper, Abdullah [1955] recognized that the remarkable feature of the tornado is not its energy but its energy density. He stated, 'The major problem in explaining the mechanics of tornadoes is to find the source of the kinetic energy and to account for its concentration over a rather small space.'

Some idea of the extraordinary energy concentration in the tornado can be obtained by estimating the temperature differences that would have to exist in the atmosphere to cause its winds. We can compute an absolute minimum temperature difference that would be required by considering the atmosphere as a perfect heat engine in which all processes are reversible and in which dissipative processes such as turbulence, entrainment, and friction do not occur. From the second law of thermodynamics the maximum work W obtainable from a heat engine operating between a source of heat at absolute temperature T_1 and a heat sink at absolute temperature T_2 is given by

$$W = Q(1 - T_2/T_1) \quad (1)$$

where Q is the amount of heat introduced at the higher temperature. In the case of the atmosphere, when we consider a mass of air that is at a temperature ΔT above the surrounding air, the heat becomes

$$Q = C_p \Delta T \quad (2)$$

where C_p is the specific heat of air at constant pressure. Consider the most convectively unstable situation possible in the normal atmosphere where T_1 is the maximum temperature of air in the troposphere and T_2 is the temperature at the tropopause.

If we assume that all the work produced in the atmospheric heat engine appears as kinetic energy of moving air, then, from equations (1) and (2), V , the maximum velocity of the air, is given by

$$V = [2C_p \Delta T(1 - T_2/T_1)]^{1/2} \quad (3)$$

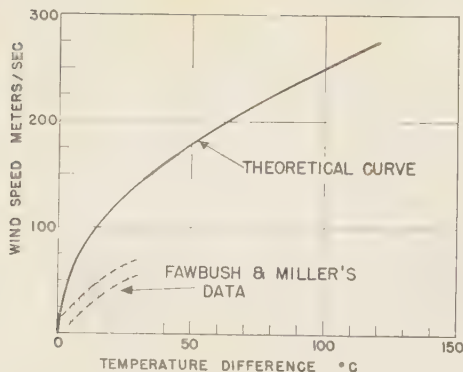


Fig. 1. Theoretical and observed winds as a function of the temperature contrast producing them.

If we substitute in this equation the value 220°K for T_2 , the heat sink, and 310°K for the heat source, then (3) becomes

$$V = 25\sqrt{\Delta T} \text{ m/sec} \quad (4)$$

Figure 1 is a plot of equation (4), showing the theoretical minimum temperature contrast between masses of air at the earth's surface that would be required to produce winds of various velocities.

Also shown in Figure 1 are the maximum gust velocities observed in thunderstorms as a function of temperature contrast as reported by *Fawbush and Miller* [1954a]. A comparison of their data with the theoretical curve shows that the actual wind speed is only of the order of one-half as large as that computed from theoretical arguments. This is to be expected, for the atmosphere is probably far from being a perfect heat engine.

From the foregoing considerations it can be seen that in order to cause the winds of near sonic velocity that appear to exist in the tornado [*Van Tassel*, 1955] temperature contrasts must occur between air masses that are well in excess of the theoretical minimum of 70°C , and probably of the order of several hundred degrees centigrade.

The puzzling fact is that none of the soundings made during conditions that cause tornadoes shows the possibility of such temperature contrasts. Even the most unstable atmospheric conditions cited by *Fawbush and Miller* [1954b] are capable of giving rise to thermal contrasts

of only 25°C between air masses. This is well below the theoretical minimum that would be required to produce tornado velocities. We can calculate that even all the heat available from the transformation of water vapor into ice in air with a dew point of 20°C could raise the temperature only 50°C .

Though it is doubtful that ordinary atmospheric processes could produce the large temperature differences required for winds of tornado speed, certain phenomena can give rise to such thermal contrasts. Quantities of heat sufficient to raise large volumes of air to a high temperature are released by volcanic activity, large fires. Tornadoes have been observed under these conditions [*Coleman*, 1946; *Hissong*, 1927; *Landsberg*, 1947; *Graham*, 1955].

The problem of understanding how thunderstorms that normally give a maximum wind of only about 50 m/sec can sometimes give tornadic winds of 200 m/sec is perplexing indeed. *Radok's* [1949] analysis of the data reported by the 'Thunderstorm Project' [*Bye*, 1948] shows that the probability of the occurrence in thunderstorms of updrafts having the velocity is only about 1 in 10^{10} . According to this the global frequency of tornadoes would not be expected to be much above 1 in 10 years.

Several investigators have sought to explain the abnormally high tornadic winds on a physical basis. *Espy* [1841] suggested that the latent heat of condensation might produce the high velocities. *Showalter* [1943] pointed out that the weight and the cooling produced by hail might cause a sufficient concentration of energy. Such mechanisms are possible, but they seem improbable. The energy available from the latent heat of condensation in a large thunderstorm cloud unquestionably is sufficient to account for the tornado, but, as we have seen, it is difficult to understand how the direct action of the condensation process could produce the high temperatures presumably necessary for the high velocities. Similarly, it is hard to see how the cooling effect of melting hail could produce the necessary temperature contrast, or how the weight of hail could produce the necessary density difference. Such a mechanism would require the unreasonably large precipitation density of several hundred grams per cubic meter.

Another possible explanation has been sug-

ed by *Brunt* [1939] and *Abdullah* [1955]. They proposed that the conservation of angular momentum causes a process whereby kinetic energy becomes concentrated in a vortex. In the foregoing calculations of the atmospheric temperature differences required to give various wind speeds, it was assumed that the work done on a mass of heated air was transformed into the kinetic energy of that same mass of air. It is possible, as the mechanism of *Brunt* and *Abdulla* suggests, that the tornado vortex may be able aerodynamically to concentrate energy imparting the equivalent of a large mass of air to a much smaller mass of air. If such a process is capable of this, it may perhaps a five or ten fold concentration in density, the preceding argument regarding the temperatures required for a tornado is not valid, and much smaller temperature differences may suffice. However, if such an aerodynamic-energy-concentrating process is possible, it does not appear to be in evidence in other rotating storms. Typhoons and hurricanes have much smaller wind velocities that are quite consistent with their temperature structure, according to the calculation.

The clue that may provide the answer to the middle of the tornado is the very intense electrical display that is commonly observed to accompany these storms. According to modern thinking [*Brooks*, 1951], this unusual electrical activity does not play a significant role in tornado dynamics. *McEachron and Jones* [1951] suggested that perhaps the tornado itself somehow causes the electrification. This appears unlikely, however, for both *Jones* and the Air Weather Service sferics observers at *Tinker Field, Oklahoma*, have shown that very intense electrical activity both precedes and follows the appearance of the funnel. If the tornado does produce the lightning that is observed to accompany it, then its energy is even more remarkable, for it will be shown that the power required for the lightning display is commensurate with the power required to produce the tornado winds themselves.

It appears to the author that the ancient idea that electricity is a cause, rather than a result, of tornadoes may afford a better explanation of the facts. The vivid electrical displays observed, particularly with nocturnal tornadoes, are so striking that the obvious deduction that tor-

nadoes might be caused by electricity is old indeed.

References to this idea can be found in the works of *Lucretius* [60 B.C.] and *Francis Bacon* [1597]. Two prominent scientists, *Peltier* [1840] in France and *Hare* [1837] in America, independently advanced similar theories in the late 1830's. *Peltier* wrote, 'Everything proves that the tornado is nothing else than a conductor formed of the clouds, which serves as a passage for a continual discharge of electricity . . .' In the same vein, *Hare* said, 'After maturely considering all the facts, I am led to suggest that a tornado is the effect of an electrified current of air superseding the more usual means of discharge between the earth and clouds, in those vivid sparks which we call lightning.'

It appears that the present eclipse of these ideas may be traced in part to objections raised by *Oersted* [1839] and *Finley* [1887]. Although as recently as 1890 we find *Hazen* [1890] calling for a re-examination of the electrical theory, it has been almost completely forgotten.

In the author's opinion, the objections raised in the past against the electrical theory are not tenable in the light of what is known today. It is the purpose of this paper to restate the case for the electrical theory of tornadoes in terms of our present knowledge of physics and meteorology. It will be shown that observations of tornadoes indicate the presence of sufficient electrical energy to account for their violence, and that there are several mechanisms by which this electrical energy could produce the intense winds.

PRESENCE OF ENERGETIC ELECTRICAL PHENOMENA IN TORNADOES

If the high velocity of the tornado is to be explained on the basis of thunderstorm electrical energy, one must first prove that intense electrical activity is closely associated with tornadoes. That this is the case is clearly evident from many different accounts that have been written about tornadoes. *Flora* [1953] quotes *Harrison's* conclusion 'that lightning of a peculiar and intense character is an almost invariable accompaniment of tornadoes,' and that airplane pilots frequently so identify tornado clouds.

The visual evidence of the close association between lightning and tornadoes is confirmed by

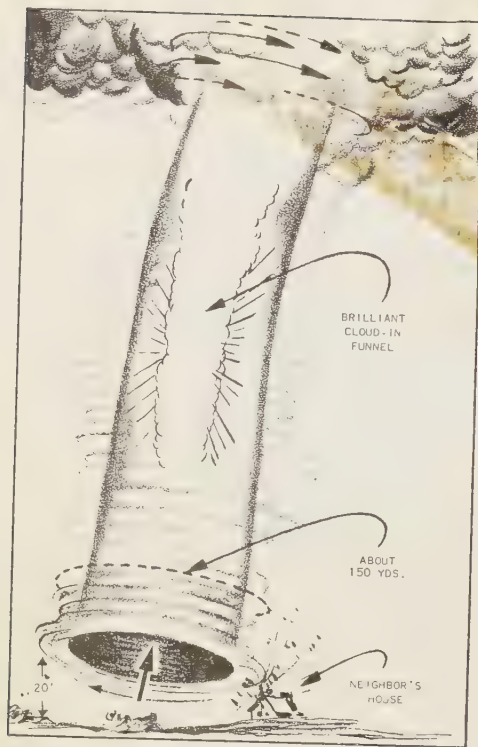


Fig. 2. Appearance of a Texas tornado according to Hall [1951] (*Weatherwise, American Meteorological Society*).

the radio 'static' measurements made by Jones [1951]. He concluded from sferics measurements that in tornado-producing storms lightning discharges occur at the rate of 10 or 20 per second, which is about ten times the rate in ordinary storms.

Further evidence of the close association between electricity and tornadoes is contained in reports of those who have looked up into the interior of a funnel and lived to tell about it. These observers report a variety of electrical phenomena, such as incessant lightning [Justice, 1930], a brilliant luminous cloud [Hall, 1951], a ball of fire [Bowker, 1953] or a display "like a Fourth of July pinwheel" [Montgomery, 1955], in the tornado tube.

Although no one has as yet photographed these striking displays, two different observers have made drawings of what they saw. Hall

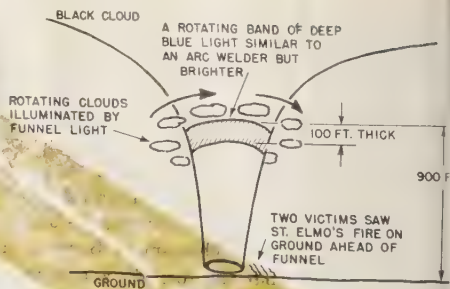


Fig. 3. Appearance of an Oklahoma tornado according to Montgomery [1956].

[1951] observed a tornado in Texas at close range and published the sketch shown as Figure 2. Montgomery [1956] watched the devastating Blackwell, Oklahoma, tornado of 1955 from a distance of about half a mile and prepared the drawing shown as Figure 3. The approximate dimensions indicated were computed from angles subtended by the tornado relative to structures near the point of observation. There are other phenomena suggestive of electrical effects. Intense St. Elmo's fire is frequently observed near the funnel, and odors, probably of ozone and nitrogen oxides, have been described. Buzzing and hissing noises suggestive of electrical discharges have been reported near the funnel. After its passage, dehydration of vegetation and the surface soil has been noted along the path.

In addition to the fairly well understood primary and secondary electrical effects discussed above, accounts of tornadoes rather frequently include mention of 'beaded lightning' and glowing or exploding fireballs [Flammarion, 1873]. These phenomena are apparently the same as or are closely related to the controversial 'ball lightning' [Goodlet, 1937; Kapitza, 1955] whose existence and nature are still debated. In view of our present almost complete ignorance, we shall make no attempt to discuss this class of observations. It is worth remarking, however, that an understanding of ball lightning may very well be necessary if the tornado puzzle is to be solved.

ESTIMATES OF AVAILABLE ELECTRICAL ENERGY

If thunderstorm electricity is the source of the tornado's concentrated energy, it must be demonstrated that there is sufficient electrical en-

available to account for the energy observed in the tornado. From the rough estimate that can be made at the present time, it appears that there is ample electrical energy.

Unfortunately, we have no knowledge of the voltage and the amounts of charge transferred in the lightning strokes associated with tornadoes. Therefore, in estimating the energy of these strokes, we must use those measurements that have been made of the lightning in ordinary storms. *Flora* [1953] stated, "Jones determined from extensive investigations that discharges of lightning in connection with tornadoes are from exceedingly high-voltage sources and are brighter, fiercer, and more vicious than those from any other type of storm." On the basis of observations such as this and the fact that the storms producing tornadoes are very large and intense, they are probably justified in inferring that the lightning strokes accompanying tornadoes are at least as energetic as those that have been measured in common storms. *Schonland* [1953] estimated that an average lightning discharge involves the dissipation of at least 6×10^{10} ergs. An estimate of the electrical power available in a tornado-producing storm can be made if we multiply the energy in each stroke by the number of strokes per second. If we assume that the 10 or 20 strokes per second observed by *Jones* [1955] are ordinary lightning strokes, then, using *Schonland's* data for the energy per stroke, we can compute that 10^{12} ergs/sec or 10^6 kw of electrical energy is being produced. In other words, rough estimates made on the basis of the meager data available indicate that there is enough electrical energy to power a tornado.

MECHANISMS BY WHICH ELECTRICAL ENERGY CAN PRODUCE WIND

It has been shown that there is very probably sufficient electrical energy in a violent thunderstorm, if properly applied, to cause a tornado. It must now be established that this electrical energy can be transformed into the kinetic energy of rapidly moving air.

There appear to be two different ways in which this might happen. The first, suggested in the comments of *Lucretius* [60 B.C.], is that electrical energy could heat a volume of air to such a high temperature that abnormally intense convection would take place. The second, pro-

posed by *Hare* [1837], is that highly charged air could be accelerated to a high velocity in a strong electric field.

Both of these processes arise through the transport of electric charge through a potential difference. In an electric storm, the flow of charge can take place in several different ways. The relative importance of the mechanisms by which electrical energy is converted into wind probably depends on the character of the electrical currents flowing in the storm. Before examining the energy conversion processes in detail let us first consider the charge transfer mechanisms that might operate in a storm.

The first and most familiar of these is the lightning stroke in which a narrow path through the air suddenly becomes highly ionized and conducting; the charge flows in a fraction of a second [*Hagenguth*, 1951]. This is an intermittent process in which the instantaneous currents reach many kiloamperes, though the average current is much less.

The second mechanism by which current may flow and heat the air is by a continuous arc or glow discharge. In this case, the current flow is steady and does not reach the high instantaneous values of the spark of lightning discharge.

The third mechanism by which current can flow is by the movement of charged particles, such as ions or electrified cloud particles, in an electric field. Such conduction processes, although usually invisible and unspectacular, probably account for the largest part of the charge transport from the earth to the clouds in thunderstorms [*Wormell*, 1930].

In all three processes a large portion of the electrical energy is released as heat. This process is most clearly evident in the case of the spark or the arc, but it is none the less real in non-luminous conduction currents. A single ordinary lightning stroke releases enough thermal energy to increase by 13°C the temperature of a column of air 5 km long and 10 meters in diameter.

It is suggested that each of the three mechanisms enumerated above can, under the proper circumstances, cause intense winds by the convection resulting from the hot air they produce. In the usual thunderstorm, the efficiency of the conversion of electrical energy into wind by this process is low because the heating effect of the lightning and conduction currents is widely dis-

tributed through the storm. Under these conditions, the thermal energy is so rapidly distributed in a large volume by conduction and mixing that it produces little convection. In the case of severe storms, such as those which produce tornadoes, the conditions can be quite different. If the rate of electric charging is sufficiently high, one lightning stroke may follow another before the ions produced by the first have a chance to recombine [Schonland, 1953]. In this case many lightning strokes will follow the same conductive path and produce intense heating of a relatively small volume of air. If these lightning strokes repeatedly strike through the same path for a sufficient length of time, a coherent sustained updraft of violent intensity could result.

A simple calculation shows that if ten strokes of lightning per second strike through the same path a stream of air 50 meters in diameter and moving 200 m/sec would be continuously heated by more than 100°C (see Appendix II). From equation (1) we can estimate that under these conditions the theoretical efficiency of conversion of the thermal energy released by lightning into the kinetic energy of wind might be as high as 40 per cent.

If the flow of charge by conduction, as described in the third mechanism, takes place by the movements of ions all having predominantly the same sign, then the transformation of electrical energy into kinetic energy can take place directly. If all the ions carry the same sign, as would be the case over a wide area during point discharge, then the movement of those ions, under the influence of an electric field, will transfer momentum directly to the air and cause it to accelerate. When this condition exists, the flow of electric current will both heat the air and cause it to move under the influence of electrical forces. The ratio of that portion of the electrical energy which is transformed into heat to that which is transformed into mechanical work will be the same as the ratio between the velocity of the ions relative to the air and the velocity of the air itself. Therefore, it is to be expected that the importance of the electrical acceleration of air will be greatest when the wind velocity is high and when the mobility of the charge carriers is low. This mobility will be smallest when the charge is carried predominantly on Aitken nuclei, dust, or cloud particles. This will be the case when the fast ions normally

present in the atmosphere or produced by point discharge become attached to cloud droplets or aerosol particles.

Some estimate of the possible magnitude of the electrical forces that could act in a charged cloud can be obtained in the following way. The maximum charge [Ladenburg, 1930] that can be acquired by a water droplet of radius a in an electric field E is given by

$$Q = 3Ea^2 \quad (1)$$

If the liquid water content of the cloud is denoted by L , then N , the number of particles per unit volume, is given by

$$N = L/(4/3 \pi a^3) \quad (2)$$

so the charge per unit volume is given by

$$NQ = 9EL/4\pi a \quad (3)$$

The electric force F acting on a unit volume is the product of the electric field times the charge or

$$F = 9E^2L/4\pi a \quad (4)$$

In a storm cloud, a typical value for the radius of a drop might be 20 microns, and for the liquid water content 2 g/m³. [Houghton, 1951] If such a cloud were charged in a field of 1 ESU/cm, which Gunn [1948] observed in an ordinary thunderstorm, then, according to equation (8), the force acting on a cubic centimeter of the cloud would be 0.07 dynes. This electric force is equivalent to that which would result if the air were heated 20°C above its environment. It seems possible that fields even higher than those measured by Gunn may exist in severe thunderstorms. In this case the electrical forces would be even more impressive.

PRESTER TORNADOES

Localized whirlwinds occur in a broad continuous spectrum of size and intensity ranging from miniscule eddies through dust devils and water spouts to full-fledged tornadoes. Since most of the small vortices and even some tornadoes occur in the absence of thunderstorm activity, it is certain that in these the energy from electrical sources is negligible.

The storms in which electrical energy may be significant are the intense tornadoes associated with vigorous thunderstorms. Since the term

tornado' is applied to a rather broad range of phenomena, it may be worth while to give a special name to this variety with which intense electrification is associated. An appropriate name for this subspecies is the word 'prester,' defined in the *Latin English Lexicon* [1851] as 'a fiery whirlwind that descends in the form of a pillar of fire.' Because the luminous activity referred to in this definition is frequently invisible or screened from view, it is suggested that the meaning of the term 'prester' be broadened to include all tornadoes exhibiting any or all of the following evidences of intense electrical activity: (1) luminous discharges in or near the funnel; (2) sensations of heat near the funnel; (3) scorching or dessication of vegetation near the mouth; (4) strong odors of ozone or nitrogen oxides near the funnel; (5) a buzzing noise resulting from point discharge that sounds like sferics; (6) unusually intense or frequent sferics impulses.

Since quite energetic whirlwinds can occur in the absence of electrical energy, it follows that if electricity can contribute to the energy of a tornado, it may do so in varying degrees. The prester variety of tornado may originate as an ordinary whirlwind that subsequently becomes intensified by electrical energy. Alternatively, it is possible that the vortex is initiated directly by electrical energy.

Though energetic electrification may be necessary to form a tornado of the prester type, it appears unlikely that the mere presence of adequate electrical energy is sufficient. Observations of dust devils and fire whirlwinds suggest that before vortices can form, adequate wind shear and vorticity must also be present. This is probably also true of tornadoes.

DESIRABLE FUTURE RESEARCHES

There is indisputably a very close association between unusually intense atmospheric electricity and tornadoes. There appear to be three possible explanations for this association:

1. The association is fortuitous.
2. The tornado produces the electricity.
3. The electricity produces the tornado.

It is important that investigations be undertaken to determine the relationship between these two phenomena. If tornadoes are caused by electrical mechanisms, then through better

understanding of these mechanisms it should be possible to increase the accuracy of tornado forecasts, to obtain warning of their presence, and, conceivably, to inhibit or prevent their formation.

In order to determine the cause and nature of tornadoes, it will be necessary to learn much more about them than is presently known. Experiments must be made to determine the temperature, pressure, and velocity distribution within the tornado funnel. With these data it should be possible to estimate the energy of a tornado and the forces that would be required to produce one.

In addition to a broad program of research into the meteorology and aerodynamics of tornadoes, it appears highly desirable to investigate the extraordinary electrical activity that is observed to accompany them. Such an investigation should not only answer the question whether the electrification is cause or effect but should also shed new light on the unsolved riddle of the mechanism of thunderstorm electrification.

The neglected idea that thunderstorm electricity may be the cause rather than the result of tornadoes suggests a number of desirable research efforts:

1. *Further sferics investigations.* The present tornado sferics research effort initiated by Jones should be continued on an intensified scale. Simultaneous measurements should be made over a wider range of frequencies in an effort to determine special characteristics of tornadic activity. Sferics measurements should be related to visual and photographic observations of the electric discharges and to electric field measurements made on the ground. It must be determined whether the extremely rapid sferics impulses observed to accompany tornadoes are produced by ordinary lightning discharges or by some different kind peculiar to these storms. It should also be determined whether these discharges are primarily of the intra-cloud or cloud-to-ground variety, and whether they occur at high altitudes or primarily in the lower part of the cloud.

2. *Electric field measurements.* Measurements of atmospheric electric parameters both preceding tornadoes and during them should be made both from the ground and from aircraft. According to many theories [Chalmers, 1957] the initiation of thunderstorm electrification de-

pend on electrical variables such as the fair weather electric field and the conductivity of the atmosphere. If this is the case, fair weather electrical measurements made prior to tornadoes may provide clues to their formation and may aid in forecasting their occurrence. Electric field measurement programs such as Gunn [1956] has conducted in the vicinity of tornadoes should be continued. The results of these observations should be correlated with sferics observations made at the same time. If such measurements are made with instruments having a speed of response capable of resolving individual strokes, it should be possible to understand better the physical basis of sferics activity. From measurements made at several stations, one may be able to make some estimate of the charge transferred in the lightning strokes in tornado-producing storms and the rate at which electrical energy is being released.

3. *Point discharge measurements.* Visual observations very close to the tornado funnel indicate the presence of St. Elmo's fire and intense point discharge. If measurements of the point discharge current flowing beneath the tornado can be obtained, it will be possible to determine whether the heating or space charge that is being produced is sufficient to cause the tornado. Measurements of the currents could be obtained by monitoring the direct currents flowing in telephone or power lines when the tornado passes. Preferably, such measurements might be made on wires erected for this special purpose.

Another method by which some estimate of point discharge currents might be obtained is to examine the scorching and dehydration of vegetation that is frequently produced by the tornado. By laboratory tests, it should be possible to determine the currents necessary to cause damage similar to that observed in the tornado path. By calibrating vegetation in this way, one could estimate the flux of point discharge current into the tornado.

4. *Photographic investigations.* Although it has long been recognized that tornado-producing storms are characterized by intense and unusual lightning displays, there appear to be few photographic records of this lightning. Similarly, though there are many descriptions and a few drawings of the electric activity in the tornado funnel, there are no photographs of these phe-

nomena. If the public is made aware of our present ignorance of the electrical effects accompanying tornadoes and the scientific value of photographs of tornadic electrical activity, it should be possible to enlist its aid in securing new and valuable photographic clues. If at the same time a broad photographic program can be undertaken to study the storms that produce tornadoes, it will be of great aid in relating meteorological observations to observations made with radar and sferics apparatus. Such a program should include still, time-lapse, and motion-picture photography from manned and automatic cameras on the ground, as well as aerial photography from airplanes and balloons. When the time-lapse photography is carried out at night or under conditions of low illumination the camera should be operated with the shutter open all the time to make the best use of available light and to provide a complete record of all electric discharges.

5. *Eyewitness accounts.* A valuable and presently neglected source of information concerning tornadoes is the eyewitness reports of observers near the funnels. The author and his associates visited the scenes of the 1955 Blackwell and 1956 Drumright, Oklahoma, tornadoes and have discovered a surprising number of people who had seen a variety of unusual electrical phenomena. It appears that such eyewitness reports of luminosity in the funnel, etc., very rarely appear in the scientific literature. Presumably such observations are not recorded either because the observers or those interviewing them fail to realize their importance. It is suggested that by a concerted effort to secure and tabulate various eyewitness observations it may be possible to piece together a far better description of the tornado than is now available.

Acknowledgments. The author wishes to thank Arthur D. Little, Inc., for supporting the earlier portions of this investigation and the U. S. Office of Naval Research and the U. S. Navy Bureau of Aeronautics for their later support of this work as a part of an investigation of atmospheric electricity. The author also wishes to acknowledge the invaluable stimulus and education that has been afforded by discussions with his associates in the laboratory. Particular thanks are due Kiah Maynard and Charles B. Moore, Jr., who have made invaluable contributions. In addition the author is grateful to Patricia Jones for having prepared the various charts and pictures that have been used.

APPENDIX I

Estimate of power to maintain tornado. Although no data are presently available on the probable power required to maintain a tornado, a crude estimate can be made of its order of magnitude. Let us consider a simplified model consisting of a tornado funnel of radius r . Long [58] has shown that in a vortex the vertical velocity is of the same order of magnitude as the horizontal velocity, and we will designate its velocity as V . The power to maintain this model is then the energy per unit time required to accelerate air from rest to this velocity. If the density of this air is denoted by ρ , then the mass of air m being accelerated per unit time is

$$m = \rho \pi r^2 V \quad (1)$$

The energy per unit time or the power P required is simply the kinetic energy of this mass of air moving at the velocity V , or

$$P = \rho \pi r^2 V^3 / 2 \quad (2)$$

If we assume a tornado having a horizontal or vertical velocity of $2.5 \times 10^4 \text{ cm sec}^{-1}$ and a radius of $5 \times 10^3 \text{ cm}$, we find, using equation (2), that the power required is about 10^{13} ergs/sec or 10^6 kw .

APPENDIX II

Calculation of updraft heating by lightning.

N = Number of lightning strokes per second
 E = Energy released per lightning stroke
 C_p = Heat capacity of air at constant pressure
 T = Temperature increase caused by heating from lightning
 r = Radius of updraft
 V = Velocity of updraft
 ρ = Density of air
 M = Mass of air transported in updraft per second

If we assume that the energy released by the lightning provides heating in the moving updraft

$$NE = \pi r^2 \rho V C_p \Delta T \quad (1)$$

$$\Delta T = NE / \pi r^2 V \rho C_p$$

Using $10^7 \text{ ergs/}^\circ\text{C}$ for the heat capacity of air, and 10^{-3} g/cm^3 for the density of air, and 6×10^{12} ergs for the energy of a lightning stroke, equation (1) reduces to

$$\Delta T = \frac{N \times 2 \times 10^{16}}{\pi r^2 V 10^{-3} 10^7} = \frac{2N}{r^2 V} 10^{12} ^\circ\text{C}$$

For 10 strokes per second and a radius of $2.5 \times 10^3 \text{ cm}$ and a velocity of $2 \times 10^4 \text{ cm/sec}$, this gives a temperature rise of almost 200°C .

REFERENCES

- Abdullah, A. J., Some aspects of the dynamics of tornadoes, *Monthly Weather Rev.*, 83, 85, 1955.
 Bacon, F., Natural history of winds—Extraordinary winds and sudden blasts—1622, *The Works of Francis Bacon*, Carey and Hart, Philadelphia, vol. 3, p. 449, 1844.
 Bowker, K., Today I am proud, *Osteopathic Mag.*, 40, 9, October 1953.
 Braham, R. R., Jr., The water and energy budgets of the thunderstorm and their relation to thunderstorm development, *J. Meteorol.*, 9, 237, 1952.
 Brooks, E. M., Tornadoes and related phenomena, *Compendium Meteorol.*, 1951.
 Brunt, D., *Physical and Dynamical Meteorology*, Cambridge University Press, 428 pp., 1939.
 Byers, H. R., et al, A Report on Thunderstorm Conditions Affecting Flight Operations, *U. S. Weather Bur. Tech. Paper* 7, 1948.
 Chalmers, J. A., *Atmospheric Electricity*, Pergamon Press, London, 1957.
 Coleman, S. H., *Volcanoes New and Old*, John Day Co., New York, pp. 135 and 141, 1946.
 Espy, J. P., *The Philosophy of Storms*, Little, Brown & Co., Boston, 304-373, 1841.
 Fawbush, E. J., and R. C. Miller, A basis for forecasting peak wind gusts in non-frontal thunderstorms, *Bull. Am. Meteorol. Soc.*, 35, 14-19, 1954a.
 Fawbush, E. J., and R. C. Miller, The types of air masses in which North American tornadoes form, *Bull. Am. Meteorol. Soc.*, 35, 154-165, 1954b.
 Finley, J. P., Tornadoes, *The Insurance Monitor*, New York, 169 pp., 1887.
 Flammarion, *The Atmosphere*, Harper & Brothers, New York, chap. VI, 1873.
 Flora, S. D., *Tornadoes of the United States*, University of Oklahoma Press, 194 pp., 1953.
 Goodlet, B. L., Lightning, *J. Inst. Elec. Engrs.*, London, 81, 1-56, 1937.
 Graham, H. E., Fire whirlwinds, *Bull. Am. Meteorol. Soc.*, 36, 99-103, 1955.
 Gunn, R., Electric field intensity inside natural clouds, *J. Appl. Phys.*, 19, 481-484, 1948.
 Gunn, R., Electric field intensity at the ground under active thunderstorms and tornadoes, *J. Meteorol.*, 13, 269-273, 1956.
 Hagenguth, J. H., The lightning discharge, *Compendium Meteorol.*, 1951.

- Hall, R. S., Inside a Texas tornado, *Weatherwise*, 4, no. 3, 1951.
- Hare, R., On the causes of the tornado or water-spout, *Am. J. Sci. and Arts*, 32, 153-161, 1837.
- Hazen, H. A., *The Tornado*, N.D.C., p. 80, 1890.
- Hissong, I. N., Whirlwinds at oil tank fire, San Luis Obispo, California, *Monthly Weather Rev.*, 54, 161-163, 1926.
- Houghton, H. G., On the physics of clouds and precipitation, *Compendium Meteorol.*, 1951.
- Jones, H. L., A spheric method of tornado identification and tracking, *Bull. Am. Meteorol. Soc.*, 32, 380-385, 1951.
- Jones, H. L., Research on tornado identification, *Third Quart. Prog. Rept., U. S. Signal Corps Research, Contract DA-36-039 SC 64436, File 421-PH-54-91*, 1955.
- Justice, A. A., Seeing the inside of a tornado, *Monthly Weather Rev.*, 58, 205-206, 1930.
- Kapitza, P., Concerning the nature of spherical lightning, *Comptes rend. acad. sci. URSS*, 101, 2, 1955.
- Ladenburg, R., Untersuchen Uber die Physikalischen Vorgange Bei der Sogenannten Elektrischen Gas Reinigung, *Ann. Physik*, 48, 863-897, 1930.
- Landsberg, H., Fire storms resulting from bombing conflagrations, *Bull. Am. Meteorol. Soc.*, 28, 2, 1947.
- Latin English Lexicon*, Harper & Brothers, 1851.
- Long, Robert R., Vortex motion in a viscous fluid, *J. Meteorol.*, 15, 108, 1958.
- Lucretius, C. T., *De Rerum Natura*, circa 60 B.C., Metrical Translation by William Ellery Leonard, E. P. Dutton, New York, book 6, p. 260, 1950.
- Montgomery, F. C., Tornadoes at Blackwell, Oklahoma, *Monthly Weather Rev.*, 83, 109, 1955.
- Montgomery, F. C., Some observations on the tornado at Blackwell, Oklahoma, May 2, 1955, *Weatherwise*, 9, no. 3, 1956.
- Oersted, H. C., On water spouts, *Am. J. Sci. and Arts*, 37, 1839. (This is a reprint of a translation published in Jameson's Edinburgh Journal, July 1839).
- Peltier, J. C. A., Translation by Robert Hare, *Am. J. Sci. and Arts*, 38, 73-86, 1840.
- Radok, U., A remarkable thunderstorm flight record, *Australian J. Sci. Research, Ser. A*, 2, 56, 1949.
- Schonland, B. F. J., *Atmospheric Electricity*, 2nd ed., John Wiley & Sons, New York, pp. 63 and 66, 1953.
- Showalter, A. K., The tornado—an analysis of antecedent meteorological conditions, in *Provisional Report on Tornadoes*, U. S. Weather Bureau, Washington, D. C., 1943.
- Van Tassel, E. L., The North Platte Valley tornado outbreak of June 27, 1955, *Monthly Weather Rev.*, 83, no. 11, 1955.
- Wormell, T. W., Vertical electric currents below thunderstorms and showers, *Proc. Roy. Soc. London*, 127, 567-590, 1930.

(Manuscript received April 10, 1959; revised October 27, 1959.)

The Distribution of Annual Tropical Cyclone Frequency¹

H. C. S. THOM

*Office of Climatology, U. S. Weather Bureau
Washington, D. C.*

Abstract. In preparation for studying and fitting tropical cyclone frequency distributions certain time trends in the climatological series are tested statistically and explanations of their possible cause proposed. The annual frequency of tropical cyclones and hurricanes is hypothesized to be a rare-event type of series. On the basis of physical and statistical principles it is shown how the Poisson and negative binomial distributions are adapted to fitting the climatological series under prescribed physical conditions. These distributions are shown to provide good statistical fits and therefore good estimates of probabilities of annual cyclone frequency. Where necessary, the effects of trend are removed and corresponding adjustments made in the distribution parameters.

Introduction. We were able to find only a small amount of previous work on the specific aspect of frequency analysis of tropical storms which we discuss here. Ray's paper, [1934] touched most closely on our topic although his interest was in conditional probabilities as they applied to the forecast problem. Namias, [1955] and Willett, [1955] discussed the long-term fluctuation aspect of our problem, but their interest was mainly in conditional prediction or forecasting based on the small systematic fluctuation which they believe exists in annual hurricane frequencies. Namias concentrated his efforts on the area off New England which will also be one phase of our study.

Our approach will be somewhat different from previous studies because a simple exhibition of some trend or oscillation of the storm frequencies, although of theoretical interest, is not an adequate solution to the practical prediction problem. This follows from the fact that such variations often explain such a small part of the variability in the time series that they are not of very practical use. To make the storm series more useful the large residual variability would first have to be characterized by probability distributions so that conditional probabilities or conditional predictions could be given. There are also the questions whether such systematic fluctuations really exist and have a meteorological

cause and whether the proportion of the total variability that they explain is large enough to justify the use of conditional probabilities or forecasts over simple probabilities or climatological predictions.

Our discussion here relates entirely to the series of annual tropical cyclone and hurricane frequencies for the U. S. East Coast (designated as the Atlantic Coast, New England and vicinity and the southern North Atlantic Ocean area) and Gulf Coast. The series for the latter include both the cyclones striking the Atlantic Coast and those in the New England area. We apply statistical tests for randomness to the climatological series of storm frequencies and fit frequency distributions appropriate to the data and the physical situation.

Randomness of the climatological series. The storm series considered are shown plotted on the upper diagrams of Figures 1, 2, 3, and 4 and in Table 3. There are a number of sources for such series, all of which resemble each other fairly closely. Our source for the series in the figures was U. S. Weather Bureau [1957], and for the New England series our source was the track charts of Tannehill [1956] with modifications suggested by the track charts in preparation in our Marine Section.

A brief examination of the upper diagrams of Figures 1 to 4 gives the impression that if the series are nonstationary, that is varying systematically in time, a linear trend will remove a large part of the nonstationarity.

¹ Prepared in cooperation with the Travelers Weather Research Center.

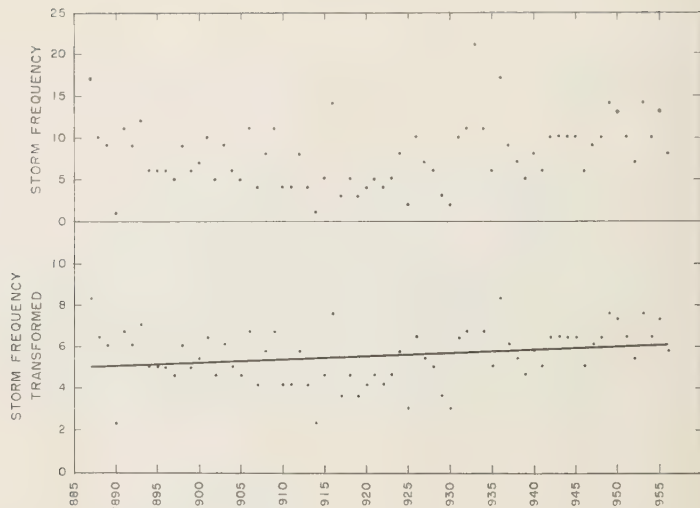


Fig. 1. Frequency of tropical cyclones, southern North Atlantic Ocean

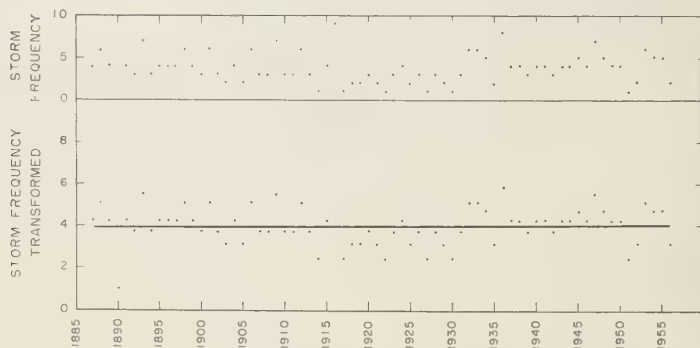


Fig. 2. Frequency of tropical cyclones reaching U. S. Atlantic Coast.

Since the most powerful tests available for trend require that the variance be uncorrelated with the mean, and this is not a property of our frequency series, we must transform the frequencies in some suitable fashion. This may be accomplished for mean frequencies greater than 1 by *Freeman and Tukey's* [1950] transformation for stabilization of Poisson variance. This is given by the equation

$$x = \sqrt{y} + \sqrt{y + 1} \quad (1)$$

As we shall see, our storm frequencies approximately follow Poisson distributions, for which the variance is proportional to the mean in

sampling. The frequencies transformed by (1) are shown on the lower diagrams of Figure 1 to 4.

The next step was to fit linear regression equations to year number as the independent variable and transformed frequency as the dependent variable. This may be conveniently done in a way which at the same time gives the analysis of variance for testing regression or trend significance and for practical appraisal of the variability explained by the linear trend. The *F*-test for frequencies of tropical cyclones and hurricanes reaching the Atlantic Coast of the United States were not significant, hence there are no

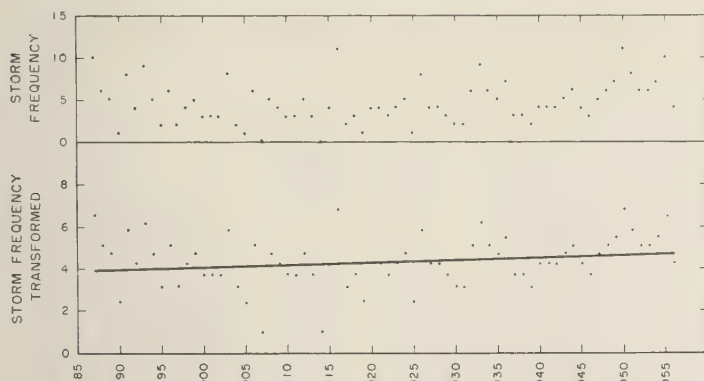


Fig. 3. Frequency of hurricanes, southern North Atlantic Ocean. *

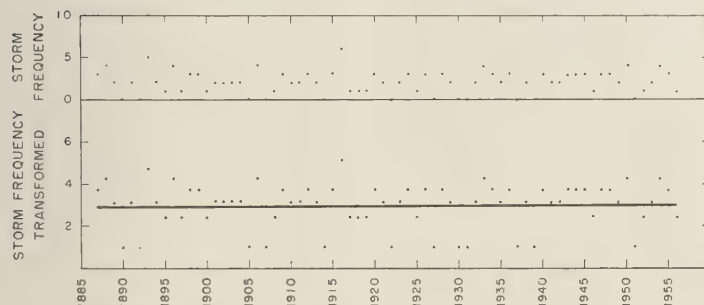


Fig. 4. Frequency of hurricanes reaching U. S. Atlantic Coast.

al trends in Figures 2 and 4 and the regression equations are of no importance. Both the tropical cyclone and the hurricane frequencies for the southern North Atlantic showed significant linear trends at the 0.10 level of probability. This we consider to be a reasonable significance level for the present situation. The analyses of variance for the hurricane frequencies are shown in tables 1 and 2.

The tests of significance give some rather striking results: Both series which involve counting the storms over the open ocean show an upward trend in annual frequency whereas both series which involve counting on land show no upward trend. The obvious conclusion to draw from this contrast is that the observational program over the oceans has gradually improved over the years and has resulted in an increase in the *observed frequency*. Over land the observational program and data through the period

TABLE 1. Tropical Cyclones, Southern North Atlantic, Analysis of Variance

	<i>S/S</i>	<i>D/F</i>	Variance
Regression	6.7585	1	6.7585
Residuals	122.9191	68	1.8076
Total	129.6776	69	1.8794

$$r^2 = 6.7585/129.6776 = 0.05212$$

$$\hat{y} = 5.0951 + 0.0154t$$

$$F = 6.7585/129.6776 = 3.74$$

of record have been quite uniform for the type of disturbance under consideration; consequently no trend in frequency occurred. This, of course, does not rule out the possibility that the observed trends might be due to an increased frequency in the more easterly areas of development. However, a simple shift in the area of development will not explain the increase, since

TABLE 2. Hurricanes, Southern North Atlantic, Analysis of Variance

	S/S	D/F	Variance
Regression	3.9794	1	3.9794
Residuals	94.3571	68	1.3876
Total	98.3365	69	1.4252

$$r^2 = 3.9794/98.3365 = 0.0405$$

$$\bar{y} = 3.9294 + 0.0118t$$

$$t' = 2.87$$

this would lead to a downward trend in cyclones crossing the coast which was not observed. Our main interest here is in the probability of tropical cyclones, so we must leave the more detailed examination of these trends to other studies.

Three more aspects of the observed trends are of importance for our purposes: First, they are not very significant statistically for it took both a fairly large, type 2 error (0.10 significance limit) and a very powerful test to detect them at all. The less powerful pair and run tests did not show a significant trend (or oscillation) in any of the series. Second, the percentage of variability explained by the trend (0.05 and 0.04) as measured by the squared correlation coefficients shown under the analysis of variance tables is small. This leaves 95 per cent or more of the variability to unexplained or random

TABLE 3. Annual Frequency, Cyclones of Tropical Origin Affecting New England

Year	No.	Year	No.	Year	No.
1901	0	1919	0	1937	0
02	2	20	0	38	1
03	1	21	0	39	1
04	2	22	0	40	1
05	0	23	1	41	0
06	0	24	1	42	0
07	0	25	0	43	0
08	0	26	0	44	1
09	0	27	1	45	1
10	0	28	0	46	1
11	0	29	1	47	0
12	0	30	0	48	0
13	0	31	0	49	0
14	0	32	1	50	0
15	0	33	1	51	0
16	1	34	2	52	1
17	0	35	0	53	2
18	1	36	1	54	3
				55	1

TABLE 4. Cyclones of Tropical Origin Affecting New England, Analysis of Variance

	S/S	D/F	Variance
Regression	1.2857	1	1.2857
Residuals	24.6679	53	.4654
Total	25.9536	54	.4800

$$r^2 = 1.2857/25.9536 = 0.0495$$

$$F = 1.2857/.4654 = 2.76$$

variability. Third, the over-all trends upward in hurricanes and tropical cyclones are about 2 a 3 in annual frequency for the 70-year record. These amount to departures of only about 1 a 1½ from the average at the extreme ends of the trend lines, which are small compared with the year-to-year variability. Nevertheless, we shall propose adjustments to the probability distribution parameters to account for these increases in frequency.

The New England series is our main interest since our final objective is to study the frequency of tropical cyclones in the New England area. Trend and frequency studies for the whole North Atlantic were made mainly to establish theory for use on the New England series. The latter series was begun in 1901 in order to avoid some of the difficulties associated with a record which could become more incomplete through extension previous to this time.

The trend problem in the New England series had to be treated differently because, although it is also a Poisson variate, its expected annual frequency is less than 1 and hence the simple transformation of equation (1) is not adequate. It was necessary to use another transformation also by *Freeman and Tukey* [1950] which gives adequate accuracy at expectations below 1. This is most easily handled as three successive approximations to be applied in order:

$$x_1 = 2(\sqrt{y+1} - \sqrt{y})$$

$$x_2 = x_1 + \frac{x_1^2 - 4}{12\sqrt{y+1}}$$

$$x_3 = x_2 + \frac{x_2 - 1}{12\sqrt{y}}$$

where x_3 is the transformed variate.

Applying this to the series in Table 3 gave

TABLE 5. Distribution of Tropical Cyclones Reaching U. S. Atlantic Coast

x	f_c	g_0	g_c	$1 - F_c$
0	0.0240	1	1.68	1.0000
1	.0895	6	6.27	.9760
2	.1669	10	11.68	.8865
3	.2075	16	14.53	.7196
4	.1935	19	13.55	.5121
5	.1444	5	10.11	.3186
6	.0898	8	6.29	.1742
7	.0478	3	3.35	.0844
8	.0223	1	1.56	.0366
9	.0092	1	.64	.0143

$$\bar{x} = 3.73; s^2 = 3.27$$

Adequacy of Poisson: $P(\chi_{90}^2 > 60.57) > 0.70$

Goodness of fit: $P(\chi_6^2 > 5.80) > 0.30$

series which could be analyzed by the analysis of variance shown in Table 4. Again, the squared correlation coefficient shows that the linear trend accounts for only about 5 per cent of the variability. The F in this instance, however, is not significant at the 0.10 probability level.

Although the trend for the New England series is not significant, we shall infer that it would be if more information were available. This inference is based on the fact that the series is a part of the tropical cyclone series for the whole North Atlantic and would therefore probably have similar properties, though perhaps to some lesser extent, because of the greater density of marine meteorological observations in the area. It will therefore also serve our purpose later to adjust the frequency distribution parameter for this trend.

The frequency distributions. The reasoning employed here in finding the physical basis for the mathematical form of the frequency distributions is similar to that used in our study of hail frequency [Thom, 1957]. The occurrence of tropical cyclones or hurricanes is a comparatively rare event; that is, the annual frequency of such events is relatively small. If the formation areas have protracted periods of favorable conditions for the formation of tropical cyclones, there will be a tendency toward correlation between successive storms or a kind of one-storm-begets-another process. Associated with this will be an increase in mean frequency. This behavior is similar to the smaller scale convection process

TABLE 6. Distribution of Hurricanes Reaching U. S. Atlantic Coast

x	f_c	g_0	g_c	$1 - F_c$
0	0.1300	12	9.10	1.0000
1	.2653	11	18.57	.8700
2	.2706	20	18.94	.6047
3	.1840	19	12.88	.3341
4	.0938	6	6.57	.1501
5	.0383	1	2.68	.0563
6	.0130	1	.91	.0180

$$\bar{x} = 2.04; s^2 = 1.84$$

Adequacy of Poisson: $P(\chi_{60}^2 > 62.10) > 0.70$

Goodness of fit: $P(\chi_3^2 > 7.44) > 0.05$

in tropical air masses on land [Thom, 1958 p. 47], where frequency and correlation are shown to be proportional.

If from such correlated series of tropical cyclones or hurricanes we select only those which reach land or those which reach the vicinity of New England, we reduce the frequency and hence the correlation or pairing, etc. This tends to make successive storms more independent in occurrence, on the average. The rare-event nature of such series and the successive independence of the events are the conditions for a Poisson distribution. Hence, we might expect that the Poisson distribution would provide a good fit to the annual frequency series for tropical cyclones reaching the United States Coast, hurricanes reaching the United States Coast, and tropical cyclones reaching the New England area. This we find to be exactly the case in Tables 5, 6, and 9a.

The correlation or pairing, etc., associated with the whole series has been shown to lead to a Polya-type process and thence to a Pascal or negative binomial distribution. This distribution would be expected to fit the series for the whole southern North Atlantic, and again the statistical tests show this to be true. This provides the basis for our analysis of the distributions.

The fitting of the Poisson and negative binomial distributions to climatological data has been discussed in detail [Thom, 1957]; consequently, we shall only indicate the parts of the computations required here. The positive binomial, Poisson, and negative binomial distributions are related as follows: For the positive

TABLE 7a. Distribution of Tropical Cyclones, Southern North Atlantic (Not Adjusted for Trend)

x	f	g_0	g_c	$1 - F$
0	0.0037	0	0.26	1.0000
1	.0151	2	1.06	.9963
2	.0347	2	2.43	.9813
3	.0587	3	4.11	.9466
4	.0816	6	5.71	.8879
5	.0986	8	6.90	.8063
6	.1072	9	7.50	.7077
7	.1072	4	7.50	.6006
8	.1002	5	7.02	.4934
9	.0887	6	6.21	.3931
10	.0748	11	5.24	.3044
11	.0607	5	4.25	.2296
12	.0476	1	3.33	.1689
13	.0362	2	2.53	.1213
14	.0268	3	1.87	.0855
15	.0194	0	1.35	.0584
16	.0137	0	.96	.0390
17	.0095	2	.67	.0253
18	.0065	0	.45	.0158
19	.0044	0	.31	.0093
20	.0029	0	.20	.0050
21	.0019	1	.13	.0021

$\bar{x} = 7.93$; $s^2 = 15.26$; $p^* = 0.9238$; $k^* = 8.58$
 Adequacy of Poisson: $P(\chi_{93}^2 > 132.8) < 0.001$
 $c = 22.0$

Goodness of fit: $P(\chi_6^2 > 9.46) > 0.10$

TABLE 7b. Distribution of Tropical Cyclones, Southern North Atlantic (Adjusted for Trend to 1947)

x	f	g_c	$1 - F$
0	0.0022	0.15	1.0000
1	.0099	0.69	.9978
2	.0248	1.74	.9879
3	.0453	3.17	.9631
4	.0673	4.71	.9178
5	.0863	6.04	.8505
6	.0991	6.94	.7642
7	.1041	7.29	.6651
8	.1018	7.13	.5610
9	.0938	6.57	.4592
10	.0823	5.76	.3654
11	.0691	4.84	.2831
12	.0559	3.91	.2140
13	.0438	3.07	.1581
14	.0333	2.33	.1143
15	.0247	1.73	.0810
16	.0179	1.25	.0563
17	.0127	.89	.0384
18	.0088	.62	.0257
19	.0060	.42	.0169
20	.0041	.29	.0109
21	.0027	.19	.0068

$\bar{x}_a = 8.56$; $s_a^2 = 14.4609$; $p_a = 0.8890$; $k_a = 9.63$

where, in our case, x is any annual frequency and μ is the mean annual frequency. To test for the adequacy of the Poisson as a proper form of distribution to fit our data series, we employ the formula [see *Thom, 1957*]

$$\chi_{n-1}^2 = \frac{n}{\sum x} \sum x^2 - \sum x \quad (6)$$

where n is the number of years of record and χ_{n-1}^2 has a chi-square distribution with $n - 1$ degrees of freedom. The best estimate of the mean is provided by

$$\bar{x} = \sum x/n \quad (7)$$

Since the negative binomial is a modification of the Poisson distribution, we may apply (6) to all frequency series to determine whether a Poisson or a negative binomial distribution should be employed. As mentioned previously, our physical hypothesis was verified, for Tables 5, 6, and 9a all show nonsignificant χ^2 s. This indicates that the Poisson distribution is suitable for fitting the series of hurricanes and tropical cyclones reaching the United States and

binomial (Bernoulli) distribution the mean and variance are

$$\mu = npq \quad (3)$$

$$\sigma^2 = npq \quad (4)$$

where $p + q = 1$. It readily follows that $p = \sigma^2/\mu$; $q = 1 - (\sigma^2/\mu)$; and $n = \mu^2/(\mu - \sigma^2)$. The relations between the mean and variance for the three distributions are as follows:

Positive binomial	$\sigma^2/\mu < 1$
Poisson	$\sigma^2/\mu = 1$
Negative binomial	$\sigma^2/\mu > 1$

It should be noted that these are not statistical criteria, for they are population values and consequently would be subject to random variation in sampling.

The Poisson distribution is a one-parameter distribution which has the probability function

$$f(x) = \mu^x e^{-\mu} / x! \quad (5)$$

TABLE 7c. Distribution of Tropical Cyclones, Southern North Atlantic (Adjusted for Trend to 1958)

x	f	g_e	$1 - F$
0	0.0016	0.11	1.0000
1	.0027	0.54	.9984
2	.0198	1.39	.9907
3	.0372	2.60	.9709
4	.0571	4.00	.9337
5	.0757	5.30	.8766
6	.0897	6.28	.8009
7	.0974	6.82	.7112
8	.0985	6.90	.6138
9	.0938	6.57	.5153
10	.0851	5.96	.4215
11	.0739	5.17	.3364
12	.0619	4.33	.2625
13	.0501	3.51	.2006
14	.0394	2.76	.1505
15	.0303	2.12	.1111
16	.0227	1.59	.0808
17	.0167	1.17	.0581
18	.0120	.84	.0414
19	.0085	.60	.0294
20	.0059	.41	.0209
21	.0041	.29	.0150

$$\bar{x}_a = 9.13; s_a^2 = 14.4609; p_a = 0.9481; k_a = 9.63$$

the series of tropical cyclones affecting New England. Tables 7a and 8a, on the other hand, both show significant χ^2 's and therefore significant divergence from the Poisson form; hence the negative binomial must be used as required by our physical hypothesis.

The negative binomial probability function is defined by the equation

$$f(x) = \frac{\Gamma(x+k)}{\Gamma(x+1)\Gamma(k)} \frac{p^x}{(1+p)^{k+x}} \quad (8)$$

where Γ indicates the gamma function, x is the annual frequency, and p and k are the parameters of the distribution. The moment estimates of these parameters are given by

$$p^* = (s^2 - \bar{x})/\bar{x} \quad (9)$$

and

$$k^* = \bar{x}^2/(s^2 - \bar{x}) \quad (10)$$

where s^2 is the moment estimate of σ^2 and \bar{x} is the sample mean as for the Poisson distribution.

The question of efficiency of estimation enters here, for the method of moments, which is

TABLE 8a. Distribution of Hurricanes, Southern North Atlantic (Not Adjusted for Trend)

x	f	g_0	g_e	$1 - F$
0	0.0202	2	1.41	1.0000
1	.0672	4	4.70	.9798
2	.1213	7	8.49	.9126
3	.1570	12	10.99	.7913
4	.1631	14	11.42	.6343
5	.1447	9	10.13	.4712
6	.1137	9	7.96	.3265
7	.0809	3	5.66	.2128
8	.0527	4	3.69	.1319
9	.0345	2	2.42	.0792
10	.0218	2	1.53	.0447
11	.0145	2	1.02	.0227

$$\bar{x} = 4.60; s^2 = 6.36; p^* = 0.3825; k^* = 12.03$$

Adequacy of Poisson distribution: $P(\chi_{68}^2 > 95.4) < 0.025$

$$C = 50.7$$

Goodness of fit: $P(\chi_s^2 > 2.01) > 0.80$

highly efficient for both the positive binomial and the Poisson, is sometimes inadequate for the negative binomial. Fisher has given a criterion [see *Thom*, 1957] for determining when the efficiency of the method of moments is satisfactory. This may be readily computed from

$$C = (1 + 1/p^*)(k^* + 2) > 20 \quad (11)$$

Fortunately, the C 's in both Tables 7a and 8a are greater than 20; hence equations (9) and (10) provide good estimates of the parameters.

Using (7), (9), and (10), we have estimated the parameters of the distributions for the data of Figures 1 to 4 and Table 3. These estimates are shown at the bottoms of distribution tables 5, 6, 7a, 8a, and 9a. Use of the estimates in (5) and (7) give the f -columns of the distribution tables corresponding to the annual frequency x . The third column g_0 is the observed number of years with x cyclones. The fourth column g_e is obtained by multiplying f by the length of record and is the expected or theoretical frequency. Comparison of g_0 and g_e gives a measure of the goodness of fit of the distribution. This comparison has been made statistically by the χ^2 test, the probabilities for which are given at the bottoms of the distribution tables. In every case the test shows that the fit is good. A comparison of g_0 and g_e in Table 9a shows a remarkable agreement between observed and

TABLE 8b. Distribution of Hurricanes, Southern North Atlantic (Adjusted for Trend to 1947)

x	f	g_c	$1 - F$
0	0.0149	1.04	1.0000
1	.0541	3.79	.9851
2	.1052	7.36	.9310
3	.1454	10.18	.8258
4	.1601	11.21	.6804
5	.1493	10.45	.5203
6	.1224	8.57	.3710
7	.0905	6.34	.2486
8	.0615	4.31	.1581
9	.0389	2.72	.0966
10	.0231	1.61	.0577
11	.0129	.09	.0346

$$\bar{x}_a = 4.89; s_a^2 = 6.1018; p_a = 0.3471; k_a = 14.09$$

calculated frequencies, an agreement seldom if ever observed in classical meteorological relationships.

Since f is a probability, it may be accumulated according to increasing x to obtain the distribution function F . When subtracted from unity this gives the fifth column $1 - F$ of the distribution tables. $1 - F$ is a more convenient form since it gives the probability that the frequency will equal or exceed x .

It may be noted that the good agreement between observed and theoretical frequencies occurs irrespective of the fact that the series fitted by the negative binomial distributions have a trend and therefore are not random. In this case the fitting is not strictly valid. In the next section we shall show the effect of this trend and adjust for it in order to obtain valid probability estimates.

Adjustment of parameters for trend. It is well known [Jeffreys, 1939] that divergence from the Poisson toward the negative binomial can be produced by two factors: the correlation or pairing mentioned above and a variation in probability from year to year. For a linear trend the effect on the Poisson distribution is easily seen. The introduction of a linear trend will not change the over-all mean but will increase the variance and hence vary the probability over years. This will cause the variance to be greater than the mean and according to our criteria will cause the distribution to tend toward a negative binomial. The analyses of variance of Tables 1 and 2 provide the means of removing the variance due to trend. Thus, the total sum

TABLE 8c. Distribution of Hurricanes, Southern North Atlantic (Adjusted for Trend to 1958)

x	f	g_c	$1 - F$
0	0.0118	0.83	1.0000
1	.0449	3.14	.9882
2	.0913	6.39	.9433
3	.1323	9.26	.8520
4	.1526	10.68	.7197
5	.1492	10.44	.5671
6	.1283	8.98	.4179
7	.0996	6.97	.2896
8	.0710	4.97	.1900
9	.0471	3.30	.1190
10	.0294	2.06	.0719
11	.0176	1.23	.0425

$$\bar{x}_a = 5.22; s_a^2 = 6.1018; p_a = 0.3705; k_a = 14.09$$

of squares (S/S) includes the variability due to linear trend and the residual S/S which would be the S/S associated with stationary random variability in the series. Hence, $1 - r^2$ multiplied by the total S/S will reduce it to the required residual S/S . If we make the reasonable assumption that the linear term explains a large part of the trend and that the correlation coefficient is nearly the same in the transformed and original data, we may remove the effect of the linear trend from the variance by multiplying by $1 - r^2$. (Our experience has shown the correlation coefficient to be almost unchanged by such a transformation.) The results are the adjusted variances shown below Tables 7b and 8b.

If we now employ the adjusted variance and the original series mean in equation (10) we obtain estimates of k from which the effect of the increase in variance or variation in probability from year to year is removed. These are shown below Tables 7b and 8b.

Another problem arises as a result of the removal of trend: Since an increase in variance due to trend produces a negative binomial tendency, a reduction in variance due to removal of trend causes a tendency toward the Poisson distribution. To test the effect of removal of trend we have adjusted the sums of squares going into equation (6) by the same factor used to adjust the variances. This reduced both χ^2 's as was to be expected: The χ^2 for the tropical cyclone series, which has the larger Polya effect,

TABLE 9a. Distribution of Cyclones of Tropical Origin Affecting New England (Not Adjusted for Trend)

x	f	g_0	g_c	$1 - F$
0	0.5902	32	32.46	1.0000
1	.3112	18	17.12	.4088
2	.0821	4	4.52	.0986
3	.0140	1	.77	.0165

$$\bar{x} = 0.527; s^2 = 0.513$$

Adequacy of Poisson: $P(\chi^2_{99} > 52.55) > 0.50$

Goodness of fit: $P(\chi^2_1 > 0.0676) > 0.70$

remained significant while the probability of the for the hurricane series was raised above the significance level, in fact, above 0.20. Since the latter is the result of a kind of pseudo test of significance, the variance of the adjusted series being still appreciably greater than the mean, we shall continue to use the negative binomial for both storm series. It will be clear that there is no definite physical boundary here between the Poisson and Polya processes.

Having removed the effect of trend, we now have a stationary series with means equal to the means at the middle of the record. Since these are not realistic values, we adjust the means to values at later dates when the storm counts are more complete or are more realistic measures of the present physical situation. In order to give samples of different situations we adjust the means to 1947, at which time one might conclude that the storm counts had attained completeness. We also adjust the series to 1958, which is meant to depict another possible aspect of complete storm count or of the present physical situation. There is the possibility, of course, that the storm count changed during the 1947-1956 period as a result of progressively increasing accuracy of measurements and observations of tropical storms. It may therefore be of interest to compare the resulting probability distributions. The adjustments to the means were made by computing the transformed values at 1947 and 1958 and converting these back to frequency by means of the regression equations shown at the bottoms of tables 1 and 2 and tables of $\sqrt{y} + \sqrt{y+1}$. There remains the problem of adjusting the estimates of the distribution parameters.

TABLE 9b. Distribution of Cyclones of Tropical Origin Affecting New England (Adjusted for Trend)

x	f	g_c	$1 - F$
0	0.4966	27.31	1.0000
1	.3476	19.12	.5034
2	.1217	6.69	.1558
3	.0284	1.56	.0341
4	.0050	.28	.0058
5	.0007	.04	.0008
6	.0001	.01	.0001

$$\bar{x} = 0.70$$

$$g = 0.4785 + 0.009631x$$

Both proposed adjustments to the parameters will be toward increases in the mean frequency. Although the parameter k varies inversely as the mean frequency and inversely as the Polya effect, the adjustments we shall make in the means are not large, and we shall assume that the k 's are constant through this adjustment. This is equivalent to assuming that the Polya effect has not been increased appreciably by the small increase in mean. Employing the adjusted variances of Tables 7b and 8b and the adjusted means of both these tables and Tables 7c and 8c in equations (9) and (10), we readily obtain the p 's and k 's for the distributions adjusted to the means at 1947 and 1958. The values of the parameters are then used in equation (8) to obtain distribution tables 7b, 7c, 8b, and 8c.

The mean frequency of cyclones of tropical origin that affect New England was adjusted to the year 1955, as this seemed to be a reasonable estimate of the present frequency in this area. The regression equation at the bottom of Table 9b and an inverse transformation based on equations (2) were used to adjust the original mean of 0.527 to a new mean of 0.70. Both these values are small enough so that the question of a Polya effect does not enter into the adjustment. Since the Poisson distribution has only one parameter, the adjustment of the mean provides all that is needed to prepare distribution table 9b.

All the adjusted tables naturally show frequencies and probabilities somewhat higher than those for the unadjusted series. For series where adjustments for trend are necessary the probability tables based on such adjustments should

always be used in applications. The $1-F$ column which gives the probability of the annual storm frequency equaling or exceeding a particular value will be most convenient in practice. For climatological work the writer would prefer the tables with statistics adjusted to 1947 and the New England table with statistics adjusted to 1955. Tables 7c and 8c, with adjustment to 1958, may be of interest in other work.

Acknowledgments. I wish to express my appreciation to Thomas F. Malone who brought this problem to my attention. For carrying out the computation of most of the tables, I am indebted to Ida Vestal, who devoted her personal time to this work. My thanks also must go to my assistant, Maurice Kasinoff, who made some supplementary computations.

REFERENCES

Freeman, M. F., and J. W. Tukey, Transformations related to the Angular and the Square

- Root, *Ann. Math. Statist.*, 21, 607-611, 1950.
 Jeffreys, H., *Theory of Probability*, Oxford Univ. Press, 67-69, 1939.
 Namias, J., Secular fluctuations in vulnerability tropical cyclones in and off New England, *Monthly Weather Rev.*, 83, 155-162, 1955.
 Ray, C. L., Tropical storm frequency and probabilities for the current season, *Bull. Am. Meteorol. Soc.*, 15, 183-184, 1934.
 Tannehill, I. R., *Hurricanes*, Princeton Univ. Press, Princeton, (1956).
 Thom, H. C. S., The frequency of hail occurrence, *Archiv für Meteorologie, Geophys. u. Bioklimatol., Ser. B.*, 8, 185-194, 1957.
 Thom, H. C. S., A method for the evaluation of hail suppression, *Z. angew. Math. u. Phys.*, 37-64, 1958.
 U. S. Weather Bureau, *Weekly Weather and Climate Bulletin, Special Summary*, 34, 3, 1957.
 Willett, H. C., A study of the tropical hurricanes along the Atlantic and Gulf Coasts of the United States, Inter-Regional Insurance Conference, New York, 1955.

(Manuscript received July 27, 1959.)

Airborne Measurement of Atmospheric Conductivity in Fifteen-Day-Old Thermonuclear Debris

ROBERT V. ANDERSON AND G. P. SERBU¹

*U. S. Naval Research Laboratory
Washington, D. C.*

Abstract. Total atmospheric conductivity was measured on a flight through fifteen-day-old thermonuclear debris. Significant increases in conductivity were noted within the debris at altitudes of 17,000 and 5000 feet. The relative increases at the two altitudes are found to be in good agreement with values calculated from ion-equilibrium conditions.

Introduction. Location and tracking of radioactive debris from nuclear tests is a problem upon which considerable effort has been expended. Measurements of radioactivity have been made at the earth's surface with both detectors and gummed paper [List, 1954; Lockhart and Baus, 1957] and aloft with filters [Machta and others, 1957]. Harris [1955] has studied the effect of radioactive debris on atmospheric conductivity near the ground and doubts that any significant effect exists at heights greater than a few meters.

Because of the availability of its research aircraft, instrumented for atmospheric electric measurements [Kraakevik and Clark, 1958], the U. S. Naval Research Laboratory was requested by the Atomic Energy Commission and the U. S. Weather Bureau to make measurements with its aircraft in radioactive debris from the Nevada tests of 1955 if and when the debris were predicted to pass within range of Washington, D. C. One such flight was made by Clark and Kraakevik on March 3, 1955, into the debris from Operation Teapot, burst 3, of March 1, 1955 [Clark and Kraakevik, 1959]. Potential gradient and both polar conductivities were measured on this flight. Within the predicted trajectory at an altitude of 20,000 feet, the measured total conductivity was some 40 per cent greater than the maximum observed in 21 other soundings. Despite this large effect on the conductivity, there was no correspondingly significant effect on either the conduction

current density or the potential gradient.

Because of these encouraging results on two-day-old debris, similar participation with the research aircraft was requested in connection with the Pacific megaton tests in 1956. Since some 48 hours' notice was required in order to stage such a flight operation, the U. S. Weather Bureau was able to furnish the U. S. Naval Research Laboratory with only an estimated trajectory based on sparse meteorological data from the Pacific Ocean, plus forecasts for at least two days. If such an estimated track fell within roughly a 1000-mile radius of Washington, D. C., an attempt at airborne detection was made.

Experimental procedure. The aircraft was equipped to measure potential gradient with a system of three electric field meters [Clark, 1957], both polar conductivities with standard Gerdien condensers [Kraakevik, 1958], and beta radiation with a Geiger counter exposed to ambient air. In all attempts, the flight was made at a constant high altitude near the aircraft ceiling in a direction roughly perpendicular to the predicted trajectory of the debris. If a maximum were noted in the total conductivity, the flight would be continued in the same direction until the conductivity had decreased appreciably from its maximum value and then, if desirable, the return flight could be made at another altitude. Since atmospheric electric conditions are so dependent on local meteorology, it was also considered desirable to conduct a vertical sounding or profile, if possible, at some time during the flight, both to determine the thickness of the exchange layer and to provide

¹ Present Affiliation: National Aeronautics and Space Administration.

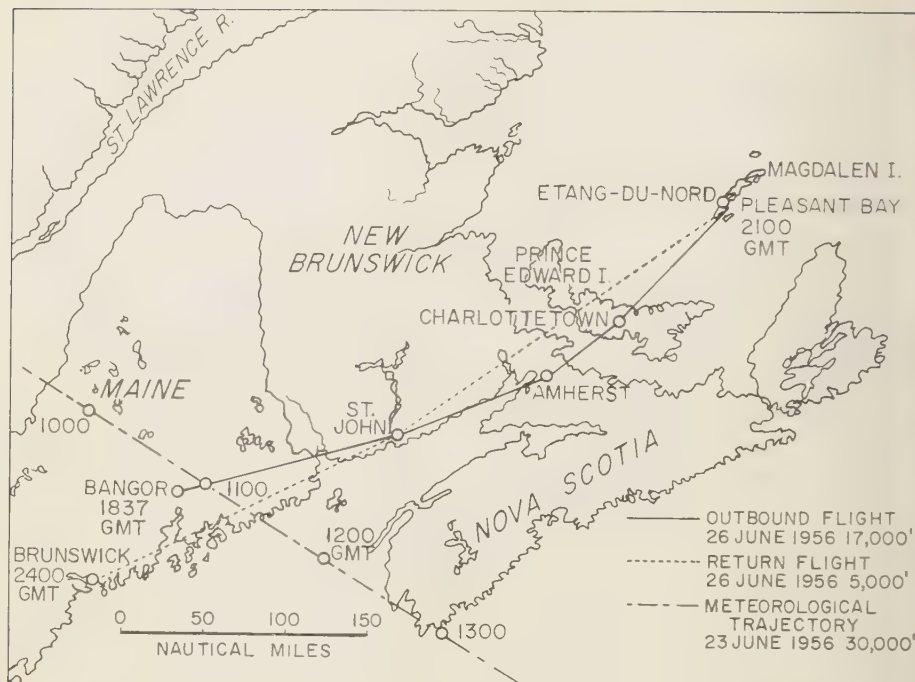


Fig. 1. Flight path of June 26, 1956.

a norm for comparison with clear-air data.

Three such attempts were made in the summer of 1956. On two of these, the actual wind vectors deviated from the conditions that had been forecast so much that the meteorological trajectory was located well to the south of the area of investigation. Consequently there were no significant results from either attempt.

The third flight was made on June 26, 1956, in an attempt to detect the debris from burst Flathead on June 11, 1956. From Washington, the aircraft was flown to Bangor, Maine, at an altitude of 17,000 feet. This transit afforded an opportunity to calibrate the instrumentation and to obtain preliminary data outside the debris area. After refueling, the flight was continued in the same northeasterly direction, as shown in Figure 1. Also shown in Figure 1 are the return flight path and a portion of the meteorological trajectory for burst Flathead. A maximum total conductivity was noted over Charlottetown, Prince Edward Island; and, since a decrease was then seen from this maximum

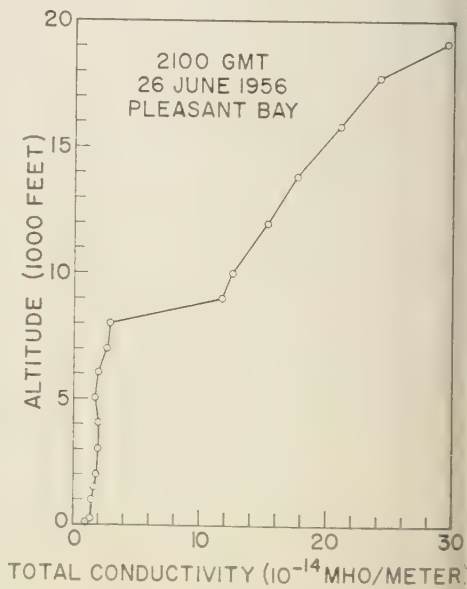


Fig. 2. Conductivity profile over Pleasant Bay.

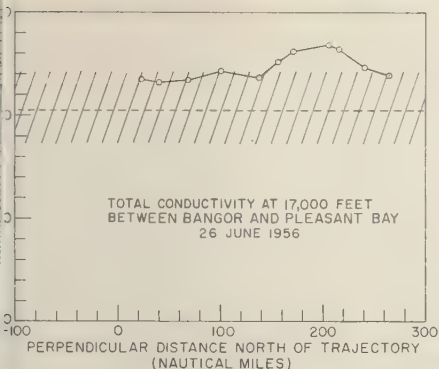


Fig. 3. Total conductivity at 17,000-ft altitude.

yond Charlottetown, a profile was made over Pleasant Bay in the Magdalen Islands. Following the profile, the return flight was made at an altitude of 5000 feet over roughly the same course and was terminated at Brunswick, Maine. **Results.** The results of the vertical sounding over Pleasant Bay are given in Figure 2, with total conductivity plotted as a function of altitude. It will be noted that there is a marked increase in conductivity between the 8000- and the 9000-foot levels. This increase has been shown to be a good indicator of the top of the exchange layer [Sagalyn and Faucher, 1954], the well-mixed, contaminated air below this level and relatively clean air above. Therefore, the 17,000-foot flight altitude was well above the exchange layer, and the 5000-foot altitude as well within it.

In Figure 3, the values of total conductivity at the 17,000-foot altitude are plotted as a function of distance from the trajectory. The shaded area denotes the range of conductivity values heretofore measured in 21 ocean flights by Kraakevik [1957], the maximum value being 2.2×10^{-14} and the minimum being 1.73×10^{-14} mho/meter. The dashed line is the average of these 21 values, 2.05×10^{-14} mho/meter. It is seen that a maximum total conductivity did occur over Charlottetown; this maximum of 2.8×10^{-14} mho/meter was some 30 per cent greater than the average and some 10 per cent greater than the maximum value previously measured. It should be further noted that the entire conductivity curve is appreciably higher than average.

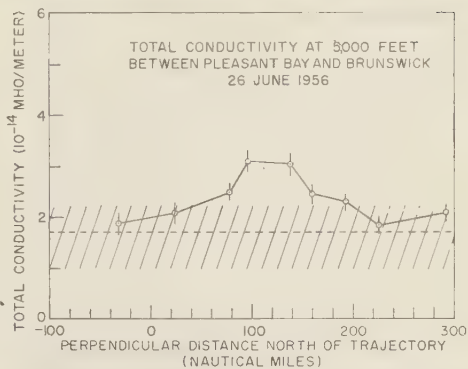


Fig. 4. Total conductivity at 5000-ft altitude.

The conductivity data from the return flight is similarly plotted in Figure 4. In considering past data, it was found that only 5 flights out of a total of 27 had an exchange layer extending above 5000 feet. Consequently, the maximum (2.22×10^{-14}), minimum (0.98×10^{-14}), and average (1.71×10^{-14}) values from only these 5 flights are indicated in Figure 4. The measured conductivity values are plotted with small lines through each point indicating the extreme variations in the measurement. It is seen that a maximum of about 3.0×10^{-14} was found between St. John and Amherst. This maximum is some 75 per cent greater than the average and some 35 per cent greater than the previously recorded maximum.

The variations in potential gradient for penetration of this debris are inconclusive, as they were in the flight of March 3, 1955, [Clark and Kraakevik, 1959]. Also inconclusive were the data from the Geiger counter, since neither the magnitude of its reading nor its variability differed significantly from background readings.

Discussion. Since the measurements were made at altitudes of 17,000 and 5000 feet, the debris particles would have taken some time to fall from the trajectory altitude of 30,000 feet. Due to a strong westerly wind at higher altitudes, the actual time of fall would be in excess of the apparent 80-hour time lag indicated in Figure 1.

Machta and others [1957] stated that most particles greater than about 10 microns will fall to the ground within a few days after a burst, whereas the fall of the remainder will be due

more to diffusion than to gravitation. Consequently, debris at higher altitudes will be trailed by a sheet of descending particles, and detection of these particles 80 hours after the passage of the parent cloud is not unreasonable.

It has been noted that the total conductivity at 5000 feet was some 75 per cent greater than the average and that at 17,000 feet was only some 30 per cent greater than the average. As a possible explanation of this, consider the ion equilibrium equation

$$q = \alpha n^2 + \beta nN \quad (1)$$

where q is the rate of ionization, α is the recombination coefficient between the small ions whose density is n , and β is the recombination coefficient between small ions and the condensation nuclei whose density is N . Above the exchange layer, where the density of nuclei is low, the second term may be neglected to a good approximation, and

$$q = \alpha n^2 \quad (2)$$

Since

$$\lambda = nek \quad (3)$$

where λ is the conductivity, k the small ion mobility, and e the electronic charge,

$$\lambda = ek\sqrt{q/\alpha} \quad (4)$$

Within the exchange layer the second term predominates, giving

$$q = \beta nN \quad (5)$$

and

$$\lambda = ekq/\beta N \quad (6)$$

By these approximations, the 30 per cent increase in conductivity at 17,000 feet would require a 70 per cent increase in q , and the 75 per cent increase in conductivity at 5000 feet would require a 75 per cent increase in the rate of ionization. It should also be noted that the background q from cosmic rays increases with altitude, so a given increase in ionization Δq due to test debris would result in a greater $\Delta q/q$ at a lower altitude.

An attempt was made to explain the more southerly maximum in conductivity at 5000 feet over that at 17,000 feet on the basis of wind shear by analyzing the weather maps for the preceding several days. However, the great speed of the cloud at 30,000 feet and inadequate meteorological data precluded any definite conclusion. The important fact is that both maxima

were found so near a trajectory which, although based on observed winds, is subject to considerable uncertainty because of the sparsity of data.

The airborne measurement of atmospheric conductivity did provide an indication of fifteen day-old thermonuclear debris. Total conductivity at 17,000 feet and at 5000 feet was found to be 30 per cent and 75 per cent greater, respectively, than the average of previously recorded values for those altitudes. The conductivity serves well in the role of an integrator with the effects of many debris particles included in a single measurement. Caution must be used, however, in any generalization of these results since the statistics are few and there is no knowledge of what effect continued nuclear testing may have on the over-all level of atmospheric conductivity.

REFERENCES

- Clark, J. F., Airborne measurement of atmospheric potential gradient, *J. Geophys. Research*, **61**, 617-628, 1957.
- Clark, J. F., and J. H. Kraakevik, Airborne atmospheric electricity measurements over a mountain ridge and in radioactive debris, *J. Meteorol.*, **1**, 1959.
- Harris, D. L., Effects of radioactive debris from nuclear explosions on the electrical conductivity of the lower atmosphere, *J. Geophys. Research*, **60**, 45-52, 1955.
- Kraakevik, J. H., The electrical conductivity and current density in the troposphere, University Microfilms, Doctoral Dissertation Ser., Pub. 21526, 1957.
- Kraakevik, J. H., The airborne measurement of atmospheric conductivity, *J. Geophys. Research*, **63**, 161-169, 1958.
- Kraakevik, J. H., and J. F. Clark, Airborne measurements of atmospheric electricity, *Trans. Am. Geophys. Union*, **39**, 827-834, 1958.
- List, R. J., On the transport of atomic debris in the atmosphere, *Bull. Am. Meteorol. Soc.*, **35**, 315-325, 1954.
- Lockhart, L. B., and R. A. Baus, Atmospheric radioactivity along the 80th Meridian, 1954, *NRL Rept. 4965*, 1957.
- Machta, L., H. L. Hamilton, Jr., L. F. Huber, R. J. List, and K. M. Nagler, Airborne measurements of atomic debris, *J. Meteorol.*, **14**, 165-175, 1957.
- Sagalyn, R. C., and G. F. Faucher, Aircraft investigation of the large-ion content and conductivity of the atmosphere and their relation to meteorological factors, *J. Atmospheric and Terrest. Phys.*, **5**, 253-272, 1954.

(Manuscript received September 4, 1959)

Sulfur in the Atmosphere

C. E. JUNGE

*Geophysics Research Directorate
Air Force Cambridge Research Center
Bedford, Massachusetts*

Abstract. In unpolluted areas sulfur occurs primarily in three compounds, SO_4^{2-} in aerosols and SO_2 and H_2S as gas. Sources of SO_4^{2-} are the ocean, the soil, and the oxidation of the gaseous compounds. A considerable fraction of the SO_2 is of anthropogenic origin. As measurements in polluted and unpolluted areas indicate, most of the H_2S seems to be natural. In unpolluted air in the northeastern part of the United States, both gases have approximately equal concentrations. The chemistry of the conversion of these gases to SO_4^{2-} is discussed.

Estimates of the sulfur budget of the global atmosphere, excluding the SO_4^{2-} in sea salt particles, indicate that 30 per cent is of anthropogenic origin. Measurements of the chemical composition of Greenland ice back to the year 1915 are presented. The measurements show that the concentration of SO_4^{2-} is much higher than that of other soluble constituents, but it has not increased since 1915 as much as one would expect if it were of industrial origin. It is concluded that industrial sulfur is removed from the atmosphere before it spreads to polar regions. This is consistent with estimates of the residence time of about 5 days for SO_2 and about 40 days for total sulfur.

Conway showed that more SO_2 enters the oceans with river water than can be accounted for by the weathering of rocks and sediments on land and that the excess must come from the atmosphere. He assumes a closed sulfur cycle and suggests that this sulfur must return to the atmosphere as H_2S , produced in the shelf areas of the oceans. In view of the considerable amounts of industrial sulfur, it is possible that the global sulfur cycle is *not* closed and that most of the excess sulfur in river waters is of anthropogenic origin and will stay in the ocean.

Introduction. Sulfur is one of the trace substances which is always found in the atmosphere, even in the most remote areas. Usually present in the gaseous state as SO_2 or H_2S , it is also an important constituent of condensation nuclei. Large amounts of sulfur are constantly being released into the atmosphere by industrial activities. To what extent the sulfur budget has been influenced by this influx during the past century is not fully understood. Except in polluted areas, however, sulfur as an important atmospheric constituent has received very little attention. The purpose of this paper is to review the subject on the basis of new observations which have become available through recent investigations in air chemistry.

The sulfur compounds in the atmosphere. Sulfur occurs in the atmosphere as SO_4^{2-} in particles and as SO_2 and H_2S . Other forms of sulfur may be present, even in unpolluted areas, but they are unknown as yet, and it is very likely that the three mentioned are by far the predominant ones.

The SO_4^{2-} in aerosols has various sources. Over

the oceans, sea water is the most important one. Extensive investigations of sea-spray aerosols have shown that sea-salt concentrations, in air, range from about 1 to 20 $\mu\text{g}/\text{m}^3$ when wind forces range between 2 and 7 Beaufort (Junge, 1958). Since the ratio of $\text{SO}_4^{2-}/\text{Cl}^-$ in sea water is 0.155, the maritime SO_4^{2-} concentrations range from about 0.10 to 2.0 $\mu\text{g}/\text{m}^3$. The concentration of SO_4^{2-} of non-maritime origin ('excess' sulfate) comprises only a small fraction over the oceans. Over land, the excess SO_4^{2-} increases rapidly to 1 to 10 $\mu\text{g}/\text{m}^3$, with maritime SO_4^{2-} dropping to negligible amounts [Junge and Werby, 1958].

The cycle of maritime SO_4^{2-} is short and of little geochemical interest, most of the SO_4^{2-} being removed over the sea by rain. In our further considerations, therefore, only the excess SO_4^{2-} will be considered. This excess SO_4^{2-} in aerosols and rain has various sources. Some fraction may come from the soil, but most likely the predominant part is formed in the atmosphere by the oxidation of SO_2 and H_2S . This is indicated by the fact that the combined concentrations of SO_2 and H_2S are, under normal

conditions, always about 1 order of magnitude higher than the excess SO_4^{2-} in aerosols. In partly polluted atmospheres, this may be as high as 2 orders of magnitude. The gaseous component, therefore, constitutes an important and always available reservoir for the generation of SO_4^{2-} in the atmosphere.

Fairly substantial data are available on SO_2 concentration in polluted atmospheres. They indicate that it varies considerably over a range of less than a hundred up to a few thousand micrograms per cubic meter. But little is known of the H_2S concentrations in polluted atmospheres. The only extensive measurements (to the knowledge of the author) are those by *Jacobs and others* [1958], who give monthly maxima, minima, and average concentrations in New York City for 1956. Compared with the SO_2 concentrations, those of H_2S are amazingly low and constant. The yearly average is $\text{max} = 5.3 \mu\text{g}/\text{m}^3$; $\text{min} = 1.7 \mu\text{g}/\text{m}^3$; $\text{average} = 3.0 \mu\text{g}/\text{m}^3$ with an over-all maximum of $7.5 \mu\text{g}/\text{m}^3$ in 249 determinations. It must be pointed out that only the *total* sulfide concentration was determined; no attempt was made to differentiate between the gaseous, liquid, or solid sulfides. It is most likely however, that H_2S is the major component.

For unpolluted, 'clean' air, separate data on SO_2 and H_2S are not available. Because of the analytical methods employed, all data obtained so far on sulfur concentration in air give more or less the combined total of SO_2 and H_2S , depending on the degree to which H_2S was oxidized during sampling. This applies also to the substantial amount of information provided by the Scandinavian network. The concentrations of total atmospheric sulfur found in this area average between 5 and $20 \mu\text{g}/\text{m}^3$, if expressed as SO_2 [*Egner and Eriksson*, 1955]. As mentioned above, the fraction of particulate sulfur which is included in this figure will be negligible. Similar concentrations of total gaseous sulfur are found by other investigators.

In order to establish the ratio of SO_2 to H_2S in unpolluted air, an attempt was made to measure these two gases separately. Air was sucked through two highly efficient bubble tubes in series, at a flow rate of $200 \text{ cm}^3/\text{sec}$. The first tube contained 30 cm^3 of a 5 per cent KClO_4 solution and a pellet of NaOH . This solution

oxidizes about 95 per cent of the SO_2 but does not affect the H_2S . The second tube contained 30 cm^3 of a 1 per cent H_2O_2 solution which oxidized all H_2S and whatever SO_2 may have passed through the first tube. The error in the determination of the ratio of the two gases is small if both concentrations are approximately equal. Both solutions were analyzed subsequently for SO_4^{2-} by the method of *Keily and Rodger* [1955].

From October through December, 1957, determinations were made at Bedford, Massachusetts, about 20 miles northeast of Boston. The air at this location, for the most part, may be considered to be representative of the New England area. In Table 1 the results of these observations are given. Although the small number of measurements do not allow any definite statistical conclusions to be made, it appears that the ratio of SO_2 to H_2S is at a minimum in air from the northwest, which is least influenced by sources of pollution. While the ratio was approximately 2 and fairly constant with northerly winds, it increased to maximum values of 10, with considerably larger fluctuations, in air from the southeast and in indifferent directions. These variations in the $\text{SO}_2/\text{H}_2\text{S}$ ratio are almost exclusively due to changes in the SO_2 concentration, whereas the H_2S concentration is rather constant and unaffected by the origin and history of the air masses (the frequency distribution of the H_2S concentrations is given in Table 2).

Along with the work of *Jacobs*, this seems to indicate that there is very little net production of H_2S in polluted areas, even in such places as New York City. The slightly higher levels of H_2S in our data may imply, on the contrary, that there is a decomposition of H_2S in polluted areas. In any case, the figures show that the H_2S is predominantly of natural origin and is a more constant atmospheric constituent than SO_2 .

The origin of SO_2 is not so easily established. The amounts of SO_2 continuously released by industrial activities are quite large, but it is not yet known whether *direct natural* sources of SO_2 are of any importance. A few measurements made in New England soil seem to confirm the belief that it is unlikely that SO_2 is released from the soil. The only other direct natural sources of SO_2 are volcanoes and forest fires, and

TABLE 1. SO₂ and H₂S Concentrations at Bedford, Massachusetts

Winds	Number of Measurements	Concentration, $\mu\text{g}/\text{m}^3$		Ratio SO ₂ /H ₂ S
		SO ₂	H ₂ S	
Southwest sector	12	17.5	8.3	2.12
Northeast sector	13	24.8	9.4	2.64
Different	7	31.1	8.9	3.50
Overall average	32	23.5	8.9	2.65

though the latter may, in many cases, be attributed to human activities. Thus far, quantitative estimations are lacking in this area.

It is possible that a fraction of the SO₂ found in the atmosphere is of *indirect* natural origin, produced by the oxidation of H₂S. There is no direct proof of this process, but a few considerations seem to point in that direction. In rural areas, for instance, the SO₄²⁻ content of rain is higher in the summertime, most likely because of a higher rate of production of H₂S at the earth's surface. It was observed that small amounts of ozone are able to oxidize H₂S to SO₄²⁻ very completely and rapidly, though this is not confirmed and is in contradiction to other experimental findings [Junge, 1958]. Finally, H₂S itself cannot easily be removed from the atmosphere because only a small fraction will be dissolved in cloud and rain droplets, and the amount absorbed at the ground surface is very small. The oxidation to SO₂ as an intermediate phase would therefore help to explain the fairly rapid removal rate of atmospheric sulfur (see section titled 'Sulfur cycle'). A direct oxidation of H₂S to SO₄²⁻ on dust and cloud particles may occur, however, and this would contradict the assumption of SO₂ as an intermediate phase unnecessary. Future studies will lead to a determination among these various possibilities. Summarizing, it can be said that most H₂S is of natural origin; that considerable amounts of SO₂, especially over land, are of anthropogenic origin; and that the extent to which SO₂ is produced directly or indirectly by natural sources is not known.

Once oxidized to SO₂, sulfur can easily be removed from the atmosphere by oxidation to SO₄²⁻, thereby transforming the gas to particulate material. Although there may be others, two processes of SO₂ oxidation are known to occur in the atmosphere: photochemical oxidation in the gas phase and catalytic oxidation in cloud droplets. According to Gerhard [1953], 0.1 to 0.2 per cent per hour of SO₂ is converted into SO₃ in bright sunlight. This, in turn, is readily transformed to H₂SO₄ in the presence of water vapor.

The oxidation of SO₂ in cloud droplets is a very efficient process [Junge and Ryan, 1957]. Only minute traces of metal ions are necessary as catalysts. This process can easily be demonstrated in the laboratory by sucking outside air, from which all particulate material is carefully removed by filtration, through a cold trap. The concentration of SO₄²⁻ found in the condensed water can only be explained by the oxidation of a substantial fraction of the SO₂ in the air. In a quantitative study of this process, it became apparent that neutralization of the SO₄²⁻ by cations was very important for its efficiency. Indeed, it is well known that SO₄²⁻, in association with NH₄⁺ or Ca⁺⁺, is a common constituent of atmospheric aerosols. Since most cloud droplets evaporate, rather than fall out as rain, the atmospheric aerosol particles will pass through several cycles of water condensation and evaporation before being removed by rain; thus SO₄²⁻ is progressively accumulated in each cycle.

The relative importance of these two processes

TABLE 2. Frequency Distribution of H₂S Concentrations

Range of concentration, $\mu\text{g}/\text{m}^3$	3-5	5-10	10-15	15-20	20
Number of measurements	10	12	6	2	2

TABLE 3. Comparison of Estimates of the Global Atmospheric Sulfur Budget

	<i>Conway</i> [1942-1943]	<i>Junge and Werby</i> [1958]
Total fuel consumption	1943: 8.0×10^8 tons/year	Average 1940-1949: 17×10^8 tons/year
Average percentage of sulfur in fuel	...	1.76
Total sulfur emission into the atmosphere	Assuming our figure 1.76% of sulfur content in fuel: 1943: 0.14×10^8 tons/year	1943: Crude oil and coal: 0.30×10^8 tons/yr Total industrial emission: 0.38×10^8 tons/yr 1956: Estimated total emission 0.40×10^8 tons/yr
Average concentration of SO_4^{2-} in rain over land over sea	3.0 mg/liter (including sea spray) ...	2.2 mg/liter (without sea spray) 0.5 mg/liter (without sea spray)
Total sulfate precipitated over land sea	3.7×10^8 tons/year ...	2.2×10^8 tons/year 1.7×10^8 tons/year
earth	3.7×10^8 tons/year	3.9×10^8 tons/year
Percentage of industrial sulfur in atmospheric sulfur budget	11 based only on sulfur in fuel	31 based on total industrial sulfur

esses in the atmosphere is hard to estimate. It is likely that the photochemical oxidation predominates in dry areas and the catalytic oxidation predominates in areas with sufficient cloud, fog, and rain formation.

Sulfur budget. The first estimations of the global sulfur budget were made by *Conway* [1942-1943], who concluded that the amount of sulfur emitted into the atmosphere by human activities is of minor importance. Recently a new estimate was made by the author. A comparison of the two estimates is given in Table 3. With the newer and better data on the total fuel consumption, sulfur emission, SO_4^{2-} rain-water concentrations, and the omission of sea-spray SO_4^{2-} , the industrial sulfur increases the global budget to approximately 30 per cent. This figure is considered to be reliable within ± 30 per cent. It indicates that air pollution, with respect to sulfur, has reached a global scale, a fact which certainly justifies a more careful consideration of the sulfur problem.

The 30 per cent increase is an average value for the entire atmosphere. As will be shown later, the average residence time of sulfur in the atmosphere is about 40 days. Data from radioactive-fallout studies indicate that the tropospheric air exchange across the equator is slow compared with this figure. Since most of the industrial sulfur is released within the Northern Hemisphere, it is realistic to restrict the estimate to the Northern Hemisphere. If the value in Table 3 for the average SO_4^{2-} concentration in rain, over land and sea, are used for the different distribution of land and sea of the Northern Hemisphere, the fraction of industrial sulfur north of the equator increases to about 50 per cent.

In this connection it may be of interest to give corresponding values for the United States alone. On the basis of a one-year analysis of rain-water samples, average data of the concentration of the total and excess SO_4^{2-} were obtained and are reproduced in Figures 1 and

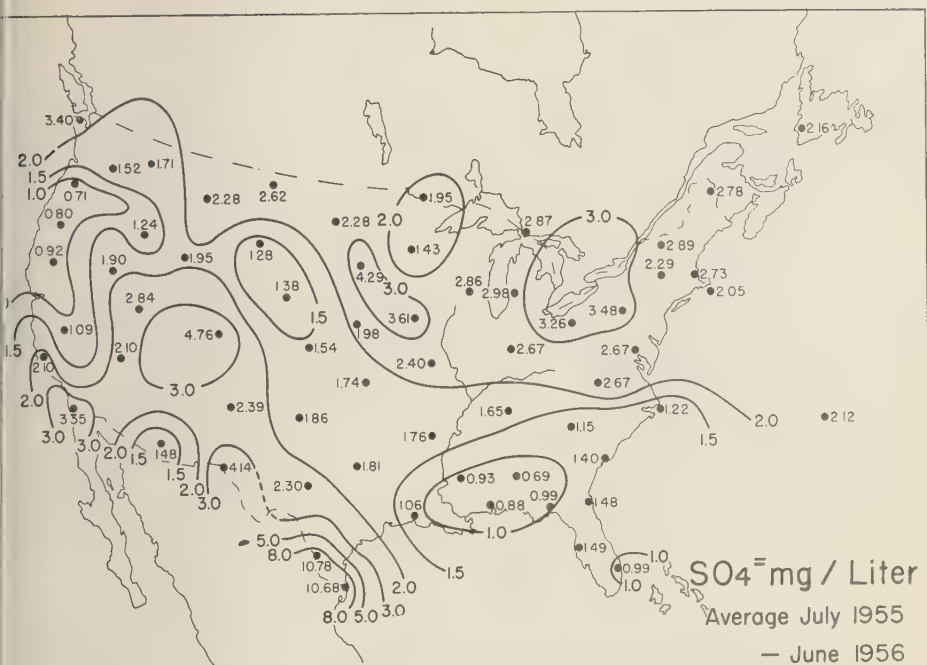


Fig. 1. Average SO₄ concentration in precipitation over the United States.

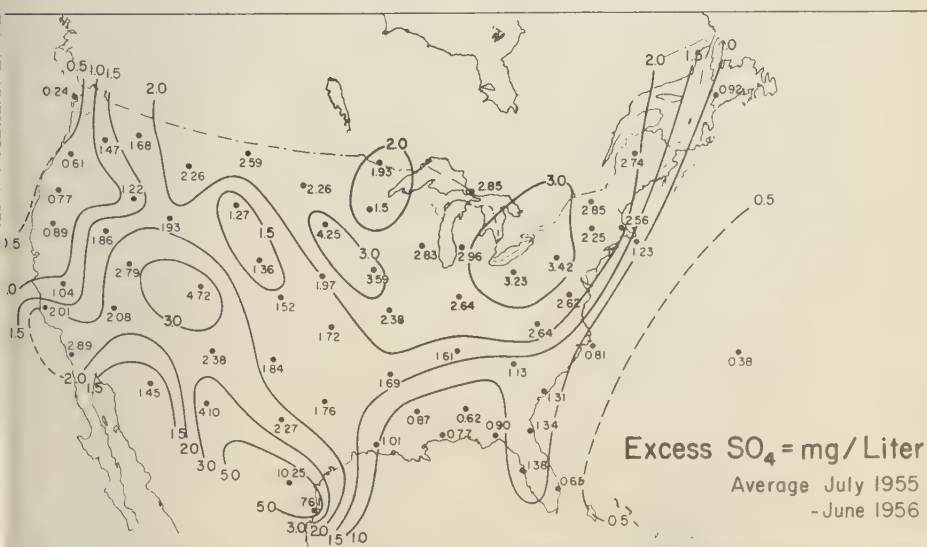


Fig. 2. Average excess SO₄ concentration in precipitation over the United States. The values in this figure are obtained from those in Figure 1 after subtraction of the maritime sulfate [Junge and Werby 58].

TABLE 4. Data on Sulfur for the United States, Calculated as SO_4^{--}

Average concentration of excess SO_4^{--} in rain	2.22 mg/l
Total amount of SO_4^{--} brought down by rain, including sea spray	1.06×10^7 tons/year
Total amount of excess SO_4^{--} brought down by rain	0.92×10^7 tons/year
Total U. S. estimated sulfur emission, 1957, calculated as SO_4^{--} *	2.59×10^7 tons/year
Total U. S. sulfur consumption, 1957, calculated as SO_4^{--} *	1.45×10^7 tons/year
Total U. S. sulfur production, 1957, calculated as SO_4^{--} *	1.88×10^7 tons/year

* Magill and others [1956].

2. The excess SO_4^{--} was obtained by subtracting the amount of SO_4^{--} which corresponded to the Cl^- analyses and the composition of sea water. With the data on the total precipitation for each station over the same period, each station having the same weight, the total SO_4^{--} removed by rain over the United States was then calculated. The results are compiled in Table 4, together with other relevant data.

Table 4 shows that the amount of total SO_4^{--} in rain is only slightly higher (about 15 per cent) than that of excess SO_4^{--} . This indicates the minor role of sea-spray SO_4^{--} over large continents. Where coastal areas are more predominant, as in Europe, the role of sea-spray SO_4^{--} is more important. The amount of total SO_4^{--} varies, of course, from year to year, within the range of normal climatological fluctuations. In the United States the total amount of SO_4^{--} found in precipitation is only 40 per cent of the total amount of industrial SO_4^{--} released. The remaining 60 per cent must escape to the oceans, primarily to the Atlantic Ocean. However, these percentages should be corrected for natural background. If we assume the background of the ocean to be 0.5 mg/l, we obtain 0.71×10^7 tons/year as a maximum of industrial SO_4^{--} precipitated over the United States. This figure gives a more realistic fraction of 72 per cent of total emitted sulfur, which escapes to the sea. This is a substantial fraction, and it indicates the degree to which air pollution in the United States influences other areas. Of interest is the fact that the total SO_4^{--} in rain over the United States is approximately equal to the total production or consumption. This gives an idea of the amounts involved in the natural sulfur cycle.

The increase of the atmospheric sulfur budget by industrial sulfur emission must have occurred during the last 50 to 100 years, during

the period of extensive industrial development. There is no information available on the global industrial sulfur emission during this period. However, since most of the sulfur is emitted in the consumption of fuel, the sulfur emission is closely parallel that of the total fuel consumption. The data of *Revelle and Suess* [1957] on total fuel consumption (Table 5) provide a general idea of the increase in sulfur emission during the same period. The question whether there is evidence of such an increase in atmospheric sulfur arises.

The only series of sulfur determinations available for this check are measurements of sulfur in precipitation made over a longer period in the United States [*Eriksson*, 1952]. In Ithaca, New York, data for the total sulfur deposited in precipitation are available for the period from 1918 to 1926. The values show a steady increase of up to nearly 100 per cent. Since Ithaca is a small city in an area which is not particularly polluted, the observed increase must be due to local sources. Another longer series of measurements, 1927 to 1942, was made at Stillwater, Oklahoma, which is in a rural area. There, a systematic trend was found. This negative

TABLE 5. Total Global Consumption of Fuel per Decade *Revelle and Suess* [1957]

Decade	Carbon Consumed 10^8 tons
1860-1869	1.5
1870-1879	2.3
1880-1889	3.5
1890-1899	5.1
1900-1909	8.2
1910-1919	11.0
1920-1929	12.8
1930-1939	13.6
1940-1949	17.4
1950-1959 est.	24.9

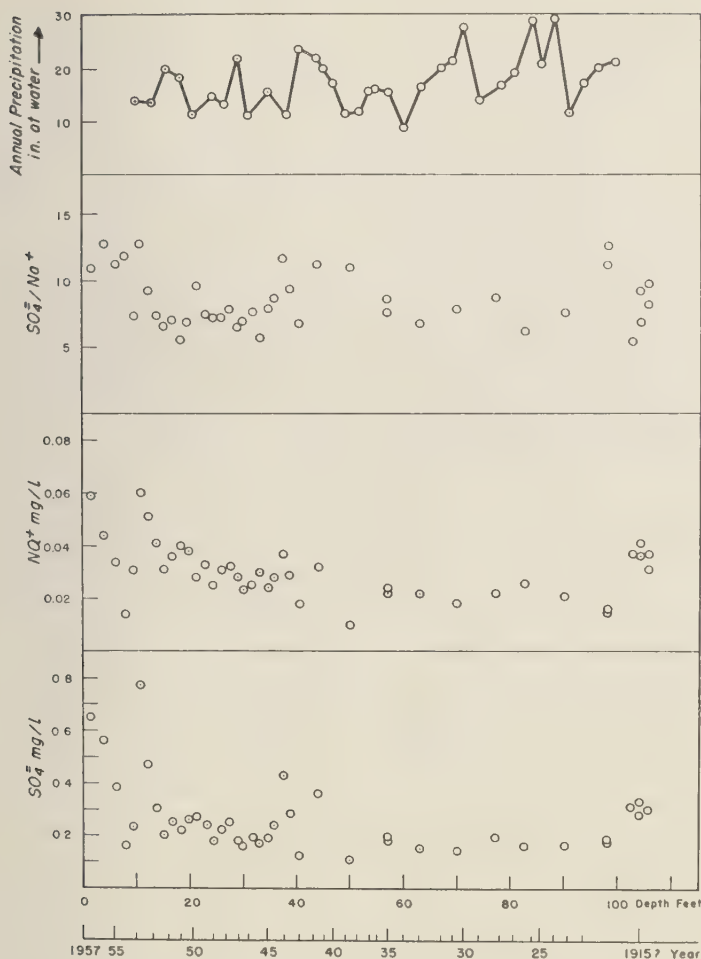


Fig. 3. Composition of ice in the center of Greenland as a function of depth or time of deposition. Data on precipitation according to *Diamond* [1956].

It is of some importance, since it cannot be due to local effects but must be representative of a wider area.

Unfortunately, all these series were much too short to allow any check to be made on the sulfur trend. In an attempt to look further into this question, ice samples from the center of Greenland were collected and analyzed in order to investigate the problem further. Precipitation on the Greenland ice cap is preserved, over the ages, in a unique way. Except during a few unusually warm summers, very little melting of

the snow occurs and the temperature is always well below freezing, falling as low as -25°C within the ice. Consequently, no percolation of melt water can occur through the ice. Because the amount of evaporation is also small, one must conclude that the original composition of the precipitation is preserved. Since general fluctuations of the precipitation regimes may result in corresponding fluctuations of the concentration of SO_4^{2-} , the ratio of SO_4^{2-} to a representative natural component, which can be considered constant, was determined. Sea spray,

represented by Na^+ , is such a component. The ice samples were collected from the northern part of central Greenland (about 200 miles east of Thule; 7600 ft high). They cover a depth range of 100 ft and the period 1915-1957. The sampling and the analyses were performed with extreme care. The results are shown in Figure 3.

It is apparent from Table 5 that the time interval between 1915 and 1957 is too short for this test, because half of the present anthropogenic sulfur production had already been reached in 1915. If spread uniformly over the Northern Hemisphere, a $\text{SO}_4^{=}$ increase of 25 per cent should have been expected since 1915, provided the increase was proportional to the natural $\text{SO}_4^{=}$ level. An increase of the $\text{SO}_4^{=}/\text{Na}^+$ ratio of this amount does not seem to be indicated in Figure 3, although the scatter of the individual values makes it difficult to say definitely that there is no increase at all. Unfortunately, earlier samples could not be obtained. The present conclusion, therefore, is that our estimates are not confirmed by the observations in Greenland.

Another interesting result of the Greenland analyses is presented in Table 6. The concentration of $\text{SO}_4^{=}$ is 1 order of magnitude higher than that of the other constituents, including sea salt. This shows that in remote places $\text{SO}_4^{=}$ is likely to be an important constituent of condensation nuclei. Some unpublished analyses from the central part of Antarctica show a similar level of $\text{SO}_4^{=}$.

The discrepancy between the estimated role of industrial sulfur and the data from Greenland can be explained in two ways: either the estimates are inaccurate or the industrial SO_2 is washed out so rapidly that no substantial fraction of it penetrates into the Arctic. The estimates are probably not in error by more than 30 per cent, but there is good evidence that SO_2 is indeed removed from the atmosphere at a faster rate than is natural sulfur. This brings us to the consideration of the sulfur cycle.

Sulfur cycle. An important parameter for the cycle of a constituent in the atmosphere is the average residence time. Basically, each of the sulfur compounds, H_2S , SO_2 , and $\text{SO}_4^{=}$, has its individual lifetime. Since there is a conver-

TABLE 6. Concentrations of Various Constituents in Greenland Ice, Site II (Averaged over the Time Period 1915-1957)

Ion	Concentration, mg/l
$\text{SO}_4^{=}$	0.25
Cl^-	0.037
Na^+	0.029
Ca^{++}	0.035
K^+	0.011

sion between the various compounds, the situation is rather complicated. With the present data, three estimates of residence times of sulfur are possible, one for total sulfur, and two for SO_2 alone.

A value for the global average for all sulfur components can be calculated by dividing the total average sulfur content in the atmosphere by the removal rate. The removal rate was determined from rain-water data to be 3.9×10^7 tons of $\text{SO}_4^{=}$ per year (Table 3). For a figure of the total sulfur content in the atmosphere, or measurements at the ground are available. If we consider primarily the values from unpolluted representative areas, such as northern Scandinavia, Florida, and Hawaii, $6 \mu\text{g}/\text{m}^3$ calculated as SO_2 appears to be reasonable average. If we assume this value to be representative for the whole troposphere, a total sulfur content of 4.6×10^7 tons is obtained. This results in a residence time of approximately 40 days. It should be noted that this lifetime stands primarily for the natural component; that is, H_2S . Because of the assumptions used for the estimation of the sulfur content of the atmosphere, the value is not very reliable.

The next two estimates are concerned with SO_2 . The rate of photochemical oxidation of SO_2 of 0.1 to 0.2 per cent per hour in normal sunlight gives an average residence time of 2 to 21 days. It should be noted that this estimate is independent of assumptions of the tropospheric content and is therefore more reliable. It represents an upper limit for SO_2 because the photochemical oxidation must go on continuously in almost all parts of the troposphere. Whatever other processes occur at the same time will shorten this value.

Another estimate of industrial SO_2 was made

Metham [1950]. He discussed the smoke (carbonaceous matter) and SO₂ distribution and transport over England. From figures for the daily production and deposit he calculated the amount which is blown out to the sea. As a result, he arrived at residence times for smoke of 6 hours and for SO₂ of 11 hours. However, a value for SO₂ was calculated by using some figures from smoke instead of SO₂ for reasons which are not apparent. By using the data for SO₂ exclusively, one obtains a residence time for SO₂ of 4 days. A similar value was given in a Leicester Report [*Metham*, 1950]. These estimates of industrial SO₂ can be considered as only tentative. In view of the fact that estimates for residence times of atmospheric aerosols are 6 days at a minimum [*Junge and Weber*, 1958], 4 days for SO₂ appears to be still rather short but more reasonable than 11 hours. We can try to estimate the residence time from the figures in Table 4 if we make simplifying assumptions. We assume that a streamline of length of x kilometers crosses the United States. Along this streamline there may be a constant SO₂ source intensity of one and a rate of removal k proportional to the SO₂ concentration (in fraction per hour). The ratio of SO₂ removed along this line to the amount released is then given by

$$\epsilon = 1 - (v/kx)[1 - \exp(-kx/v)]$$

where v is the wind velocity. For x a value of 3000 km can be assumed, which is an average between the length and the width of the United States. With an average value of $v = 30$ km/hour we obtain the values for ϵ which are given in Table 7. Our value $\epsilon = 28$ per cent corresponds to a residence time of about 7 days. Although the model used is admittedly crude and simplified, the result indicates some confirmation that the residence time for industrial SO₂ is approximately 5 days. If this figure is correct, it means that the removal of SO₂ in polluted areas is 7 times faster than the rate of photochemical oxidation. This implies that catalytic oxidation is the predominant process of removal of industrial SO₂ in temperate climates such as England and the United States have. Because of the influence of the kind and concentration of catalytic material, the removal rate will be highest for

TABLE 7. Calculation of SO₂ Removal over the United States

Assumed SO ₂ residence time in days	2	4	8	16
Corresponding value of	0.58	0.38	0.22	0.12

polluted areas and will very likely approach that for photochemical reaction in unpolluted dry climates.

The residence times of about 5 days for industrial SO₂ is short, compared with 40 days for the total, mostly natural sulfur. It is thus understandable that the natural sulfur components are found throughout the troposphere, including such remote places as the polar regions, whereas the washout of industrial sulfur will be restricted to the mid-latitudes where the major sources are located.

The removal of sulfur from the atmosphere is predominantly achieved by precipitation. This can be concluded from data on other components, including radioactive material, which indicate that dry fallout amounts to not more than about 20 per cent of the total fallout in temperate latitudes. For sulfur there is the further possibility of direct H₂S and SO₂ absorption or chemical sorption on the soil surface. These processes will tend to increase the volume of the total sulfur budget and consequently to decrease the fraction of anthropogenic sulfur. We think, however, that these effects are of minor importance.

Once the sulfur is removed from the atmosphere, the question arises as to where and how it re-enters the atmosphere. It appears that the passage through the atmosphere is linked to another interesting part of the sulfur cycle which occurs at or below the earth's surface. *Conway* [1942-1943], in his studies of the chemical evolution of the oceans, has discussed this subject in great detail. From large numbers of chemical analyses, he derived average global figures for the composition of river waters and of igneous and sedimentary rocks, and for the amounts of various elements which are dissolved during the process of weathering and are carried by the rivers into the oceans. These

figures show that the amount of material continuously added to the ocean can be accounted for by weathering, with the exception of Cl^- and SO_4^{2-} . These two ions are carried with the river water in amounts considerably in excess of those estimated for the weathering process. Since the agreement for the other compounds is very satisfactory, there can be little doubt that river water contains over 3 times as much SO_4^{2-} as can be accounted for by weathering. Conway gives the following figures for the SO_4^{2-} in river water:

Average SO_4^{2-} concentration of river water = 17.7 mg/l.
 Total volume of river-runoff = 2.4×10^{13} tons/year.
 Total SO_4^{2-} runoff therefore = 4.3×10^9 tons/year.

This is comparable to his value for total SO_4^{2-} in rain over land (Table 3) of 3.7×10^9 tons/year. Our figure of 2.2×10^9 tons/year does not include the sea spray and is therefore not comparable.

On the basis of these figures Conway is forced to conclude that most of the SO_4^{2-} carried by the rivers into the oceans can come only from the atmosphere. If it is assumed that the sulfur cycle is a closed one, the sulfur must re-enter the atmosphere over the oceans. Conway suggests that the sulfur which enters the oceans is precipitated in the shallow shelf areas which extend generally some 200 miles from the coast. Most of this area is covered by the Blue Mud, the color of which is due to organic matter and sulfides. There the SO_4^{2-} is reduced and finally converted to H_2S , which rises through the ocean water to re-enter the atmosphere. Some rough calculations indicate that this is a possible process.

This suggested cycle poses some problems, however. Figure 2 shows that the concentration of excess SO_4^{2-} is much higher over land than over sea. If most of the sulfur goes through the indicated cycle, then most of it should be removed again in the vicinity of the source areas; that is, along the coast. Measurements of inert and radioactive material in rain water indicate that the average concentration of a constituent is generally proportional to the con-

TABLE 8. Average Removal Rates of Excess SO_4^{2-} over Land and Sea

	Area, km^2	Total SO_4^{2-} Precipitated, tons/year	Rate of SO_4^{2-} Removal, tons/ km^2 yr
Land	1.48×10^8	2.2×10^8	1.49
Sea	3.61×10^8	1.7×10^8	0.47

centration in air, other conditions being equal. Figure 2 would imply, therefore, that the high concentrations of sulfur are over land and not over the shelf areas. Not only the concentration in rain water, but also the total removal rate of excess sulfur, is higher over land than over sea, as Table 8 shows.

The ratio of the removal rates, land to sea, is 3.2. This is less than the ratio of the rain-water concentrations but is still considerably larger than 1.

A more direct check could be made with data on sulfur concentration in air. Unfortunately most of the measurements of the sulfur concentrations in air are made over land, and very few data are available for comparison with coastal and ocean areas. So far, a survey of the available data seems to show that the concentrations are higher over land, even in fairly unpolluted areas, and that the values are lower over the oceans. It would be of interest to obtain better data along cross sections through unpolluted continental and coastal areas.

There are two ways in which this discrepancy can be explained. First, the assumption of a closed cycle may not be true. This is most likely the case for the industrial sulfur injected into the atmosphere. There would be no difficulty on the basis of Table 4 in explaining the excess SO_4^{2-} over the continental United States as attributable to industrial activities. If we assume this to be the case for other land areas as well, there would be no need for SO_4^{2-} from rivers to re-enter the atmosphere as H_2S .

But rain-water concentration data seem to indicate that the excess SO_4^{2-} concentration over land, even in remote unpolluted areas, is higher than over the oceans. The lowest values for the United States, Europe, and other places are 0.9 mg/l (if we disregard coastal and polluted regions), as compared with 0.5 over the ocean.

it is unlikely that all the excess continental SO_4^{2-} in rain is industrial sulfur.

As a second possibility, if we assume the same natural H_2S and SO_2 concentrations over land and ocean, the higher rate of removal of natural sulfur over land may be due to a more efficient removal process. The catalytic oxidation of SO_2 in droplets is such a process, because it depends to a high degree on the presence of metals in condensation nuclei. The omnipresence of dust in the atmosphere over land provides a ample source of such material. In other words, the higher removal rate over land would not be caused by a higher concentration of gaseous sulfur components but by a faster oxidation of SO_2 and, therefore, washout by rain. The data available are too few to allow one to decide whether the industrial SO_2 or a faster rate of removal is the predominant reason for the larger amounts of SO_4^{2-} being removed from the atmosphere over land. In the first case, a higher rate of sulfur deposition in the shelf areas, especially around industrialized continents, may have occurred during the last 50 to 100 years.

Acknowledgments. The analyses of the Bedford samples were made by T. Ryan. The analyses of the Greenland samples were made by R. Werby. Securing the ice samples from Greenland was possible only with the active support of personnel of SIPRE. All this help and support is gratefully acknowledged.

REFERENCES

- Omey, E. J., Mean geochemical data in relation to oceanic evolution, *Proc. Roy. Irish Acad. A*, **48**, 119-159, 1942-1943.
- Diamond, M., Precipitation trends in Greenland during the past 30 years, *Research Rept. 22*, Snow, Ice and Permafrost Research Establishment, Wilmette, Illinois, 1956.
- Egner, H., and E. Eriksson, Current data on the chemical composition of air and precipitation, *Tellus*, **7**, 134-139 and subsequent issues, 1955.
- Eriksson, E., Composition of atmospheric precipitation II, Sulfur, chloride, iodine compounds, *Tellus*, **4**, 280-303, 1952.
- Gerhard, E. R., The photochemical oxidation of sulphur dioxide to sulphur trioxide and its effect on fog formation, *Tech. Rept. 1*, Contract SF-9, Eng. Expt. Sta., Univ. of Illinois, Urbana, Illinois, 101 pp., 1953.
- Jacobs, M. B., M. M. Braverman, and S. Hochheiser, Ultramicrodetermination of sulfides in air, *Anal. Chem.*, **29**, 1349-1351, 1958.
- Junge, C. E., Atmospheric chemistry, in *Advances in Geophysics*, **4**, 1-108, Academic Press, Inc., New York, 1958.
- Junge, C. E., and T. Ryan, The oxidation of sulfur dioxide in dilute solutions, *Quart. J. Roy. Meteorol. Soc.*, **84**, 46-55, 1957.
- Junge, C. E., and R. T. Werby, The concentration of chloride, sodium, potassium, calcium and sulfate, *J. Meteorol.*, **15**, 417-425, 1958.
- Keily, H. J., and L. B. Rodgers, Nephelometric determinations of sulfate impurities in certain reagent grade salts, *Anal. Chem.*, **27**, 759-762, 1955.
- Magill, P. L., F. R. Holden, and C. Ackley (editors), *Air Pollution Handbook*, McGraw-Hill, New York, 2-45, 1956.
- Metham, A. R., Natural removal of pollution from the atmosphere, *Quart. J. Roy. Meteorol. Soc.*, **76**, 359-371, 1950.
- Revelle, R., and H. Suess, Carbon dioxide exchange between atmosphere and ocean and the question of an increase of atmospheric CO_2 during the past decades, *Tellus*, **9**, 18-27, 1957.

(Manuscript received June 19, 1959; revised October 6, 1959.)

On the Dynamical Structure of the Gulf Stream as an Equivalent-Barotropic Flow

G. NEUMANN

*Department of Meteorology and Oceanography
New York University, New York 53, N. Y.*

Abstract. Some of the main features of the Gulf Stream and its horizontal velocity profile can be accounted for by treating the stream as a nonaccelerated, equivalent-barotropic flow, determined by pressure, Coriolis, and frictional forces. An essential part of the model is an inclined lower boundary of the Stream in which the depth increases with distance from the coast to the Sargasso Sea. Horizontal profiles of the average velocity for vertical columns of water in the Gulf Stream are computed for different assumptions about the form of the internal friction. They verify a sharply peaked distribution of velocity across the Stream and an asymmetrical horizontal shear. The cyclonic shear region near the coast is stronger than the anticyclonic shear region toward the Sargasso Sea.

Introduction. Of all the ocean currents, the Gulf Stream has been and still remains the most challenging to be understood and to be explained. Systematic observations conducted by the Woods Hole Oceanographic Institution have revealed remarkable details of the flow pattern, especially in the surface layers. Not only has this work been of extreme help toward a better phenomenological understanding of the Stream, but it has also inspired interesting analytical studies. Existing theories of the oceanic circulation attempt to account for the established features of this swiftest of all currents. The general westward intensification of the wind-driven circulation in an ocean of constant depth was explained by *Stommel* [1948] as being the result of the variation of the Coriolis parameter with latitude. Subsequent work by *Hidaka* [1949], *Frank* [1950], *Hansen* [1951], *Hassan* [1958], and others, who used more refined theoretical models, contributed to a further understanding of the large-scale flow in the oceans, and these investigators also showed that the wind-driven, horizontal circulation, in general, and the Gulf Stream, in particular, need further study.

In 1955 it was suggested that the fact that the depth of the wind-driven circulation is not constant but varies considerably, especially in the meridional direction throughout the oceans, is an important feature to study [*Neumann*, 1955]. For a homogeneous ocean, where the lower boundary of the currents is the bottom,

it was shown by *Ekman* [1923] that a direct influence of the bottom topography on the circulation system exists. In a baroclinic ocean, where the currents do not reach the bottom but nevertheless vary in depth, the effect of the topography of the lower boundary on the circulation may not be so obvious.

The oceans are stratified in density in such a way that the density generally increases with depth. This is the reason why, in most parts of the ocean, stationary wind-driven currents, including slope currents, decrease with depth and become zero before they reach the bottom. The slope of the sea surface and the distribution of mass balance in such a way that, at some depth above the bottom, the horizontal pressure gradient vanishes. If this level represents a level of no absolute motion, it can be considered to be replaced by a solid boundary. Certain relationships have to be fulfilled between the circulation and its lower boundary since they are not independent of each other. This effect is the same as the effect of the true bottom on currents in a homogeneous ocean.

The purpose of the present article is to show that certain main features of the Gulf Stream and its counterpart in the Pacific Ocean, the Kuro Shio, can be accounted for by treating the current as an unaccelerated flow, determined by the balance of pressure, Coriolis, and frictional forces if the depth of the current is considered in the equations for the balance. This is in con-

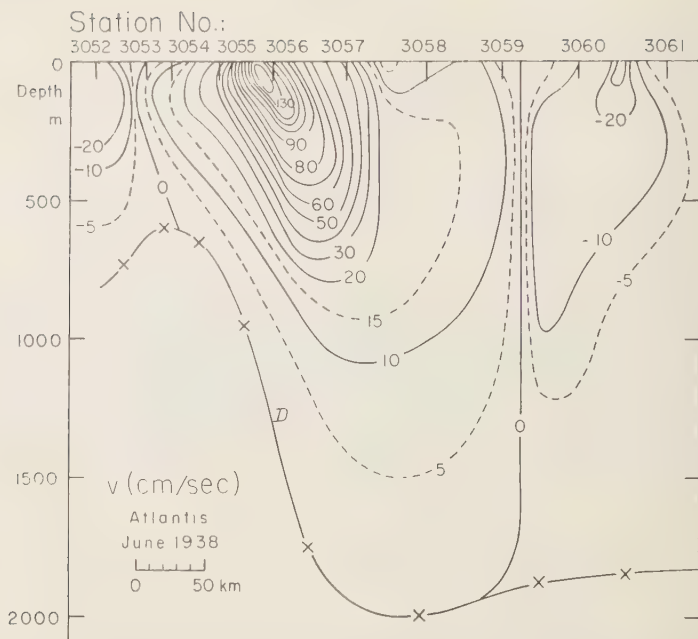


Fig. 1. Velocity distribution in a cross section through the Gulf Stream computed from *Atlantis* observations (Montauk Pt.-Bermuda section, June 1938). The line -x-x- represents the lower boundary *D* of the current (reference level) [from Neumann, 1956].

trast to a model of the Gulf Stream given by Charney [1955], who disregarded frictional forces but invoked nonlinear field accelerations. A model similar to that of Charney was recently adopted by Defant [1957] with the difference that instead of a two-layer model, the concept of an equivalent-barotropic flow [Neumann, 1956] was introduced. Both Charney and Defant explained the anticyclonic shear region on the side of the Gulf Stream toward the Sargasso Sea. The peak of the horizontal velocity profile in the main axis of the Gulf Stream and the sharp decrease of the velocities in the cyclonic shear region toward the continental shelf remained unexplained.

Some facts about the Gulf Stream. An essential part of the Gulf Stream is its intense baroclinity. Cross sections of density across the Stream reveal the strongest slope of isopycnal surfaces known in the oceans. The density increases by as much as 2 units of σ_t , or more, in the upper 500-meter layer between the eastern and western edge of the current.

Another essential fact is that in its southern

part the Gulf Stream flows partly over the continental slope and is guided by it. In this region the Stream reaches the ocean bottom and its flow is directly influenced by the topography of the continental slope. The Stream is deep along its eastern edge and shallow along its western edge. After separation from the continental shelf, approximately off Cape Hatteras, the Gulf Stream tends to maintain a free lower boundary along the eastern edge at a depth of about 1800 to 2000 meters, whereas along its western edge it is only 500 to 700 meters deep [Defant, 1941; Neumann, 1956].

The shape of the lower boundary, *D* of the Gulf Stream, after it leaves Cape Hatteras, is shown in Figure 1 for the *Atlantis* section at station numbers 3052 to 3061 at the top of the figure. *D* is computed from the mass distribution and the relative horizontal pressure field according to Defant's method [1941]. Although this method is based essentially on hypothetical ideas and does not always permit unique conclusions, it seems to work well, especially in regions of strong stratification. Direct current

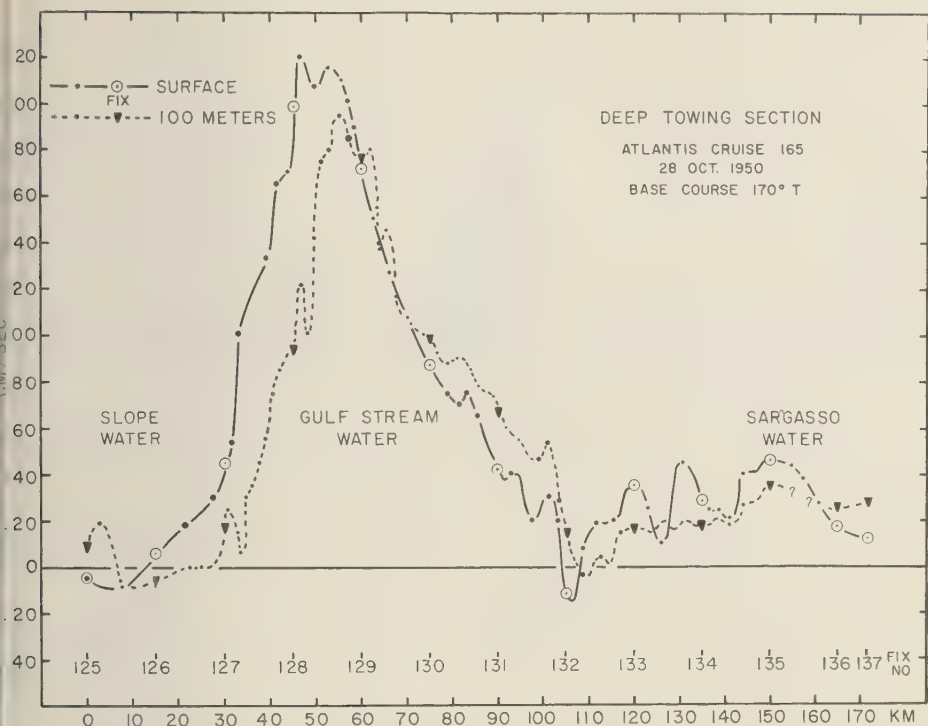


Fig. 2. Results of simultaneous electromagnetic measurements (G.E.K.) of the velocity at the surface and at 100-meter levels in the Gulf Stream [from von Arx, 1952].

measurements in the anticyclonic shear region recently carried out by Woods Hole Oceanographic Institution in cooperation with the National Institute of Oceanography, Wormley, England [Swallow and Worthington, 1957], have confirmed the fact that at least in this part of the Gulf Stream the lower boundary is found between 1000 and 2000 meters. Below this depth the current reverses its direction and flows with a speed of 5 to 10 cm/sec toward the south, as was also derived by Defant [1941]. It would be of great value to continue such measurements across the Gulf Stream and to determine its depth along the continental side.

If a lower boundary of the Gulf Stream, as shown in Figure 1, is accepted, another important effect about the density distribution $\rho(x, y, z)$ in the moving layer can be obtained. Although the horizontal density gradient in the individual level surfaces across the Stream is extremely strong, the mean density for vertical columns of water between the sea surface ζ and the lower

boundary D (equation 2) is practically constant. This has been proved for a great number of sections across the Gulf Stream in different regions [Neumann, 1956; Martineau, 1958]. It seems that the density field of the Gulf Stream tends to approach a system that may be called equivalent-barotropic. Such a system not only requires a lower boundary of variable depth D that slopes downward from the continental side to the Sargasso Sea, but it also requires a main current axis (maximum velocity) which gradually shifts to the eastern part of the Stream with increasing depth (Fig. 1). This feature was indicated many years ago by Iselin [1936] in the Atlantis section 1225-1230 (April 20-22, 1932) and has since been found in most velocity sections of the Gulf Stream and the Kuro Shio by means of the classical method of dynamical computations. However, it was neither discussed nor explained.

The results obtained by von Arx [1952] have revealed more details of the horizontal velocity

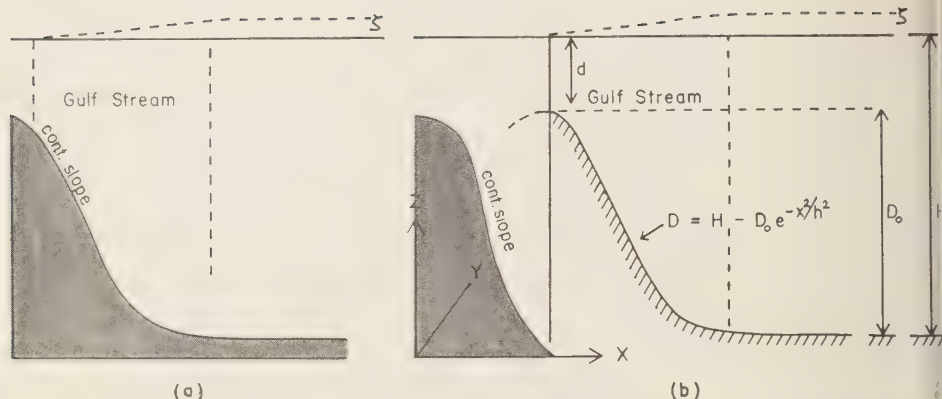


Fig. 3. Schematic representation of the location of the Gulf Stream in its southern part, where it flows over the continental slope (Fig. 3a), and in its northern part after separation from the continental slope (Fig. 3b). The line D represents the free lower boundary of the Stream.

distribution in the Gulf Stream. Figure 2 is an example of simultaneous measurements of the velocity across the Gulf Stream off Cape May. Although the figure presents the uncorrected measurements, it serves to illustrate three important facts: a sharply peaked distribution of velocity across the current, an asymmetrical horizontal shear on either side of the current maximum, and a displacement to the right of the profile at 100 meters.

The model and its basic equation. An essential requirement of the model to be introduced is that the lower boundary of the Stream be inclined in such a way that the moving layer is shallow (about 500 meters) along the continental side and about 2000 meters deep near the Sargasso Sea. As long as the Stream flows over the continental slope, the lower boundary is the bottom, as indicated in Figure 3a. From the point of separation northward, the Stream is assumed to maintain the general inclination of D as shown in Figure 3b. The depth distribution is approximated by

$$D = H - D_0 \exp(-x^2/h^2) \quad (1)$$

where D_0 is 1500 meters, $H = 2000$ meters, and $d = H - D_0 = 500$ meters is the depth at the western edge of the Stream. The x axis points to the east, the y axis to the north, and z is counted positively upward from undisturbed sea level. A length parameter h of 100 km is introduced to make the horizontal distance dimensionless.

On account of the equivalent-barotropicity of the Gulf Stream, $\partial \bar{p} / \partial x = 0$, where \bar{p} is defined by

$$\bar{p}(x, y) = \frac{1}{\xi - D} \int_D^\xi \rho(z) dz \quad (2)$$

There is much evidence that neither \bar{p} nor D changes in the direction of the flow. Whether or not this is exactly true cannot be decided at the present time. If there are small changes of \bar{p} and D in the flow direction, they are certainly so small that they cannot be detected by present means [Martineau, 1958]. In the model of this analysis, the assumptions are made that $\partial \bar{p} / \partial y = 0$ and $\partial D / \partial y = 0$, whereas $\partial D / \partial x \neq 0$.

The general equations of motion for a non-accelerated flow driven by the wind can be written

$$\left. \begin{aligned} f\rho v + \mu \frac{\partial^2 u}{\partial z^2} - \rho u v - \frac{\partial p}{\partial x} &= 0 \\ -f\rho u + \mu \frac{\partial^2 v}{\partial z^2} - \rho v^2 - \frac{\partial p}{\partial y} &= 0 \end{aligned} \right\} \quad (3)$$

where $f = 2\omega \sin \phi$, and where u and v are the horizontal velocity components in the x and y directions, respectively. The pressure $p(x, y, z)$ depends on the slope of the sea surface $\xi(x, y)$ and the distribution of density $\rho(x, y, z)$. The form for the frictional forces in (3) has been chosen, following Stommel [1948], in order to make the analysis simpler. Frictional forces resulting from vertical shear are separated from

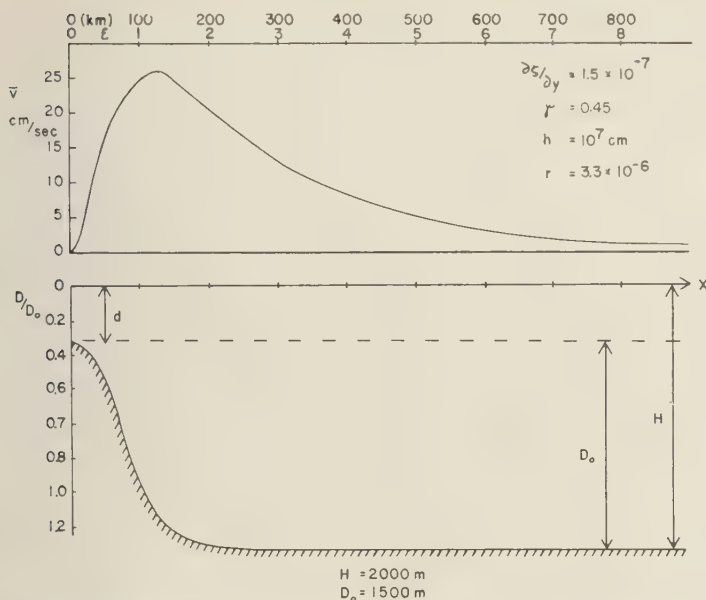


Fig. 4. Average velocity \bar{v} of vertical columns of water between the inclined lower boundary D of a meridional current and the sea surface (computed for $\gamma = \beta h/r = 0.45$). The profile of the lower boundary across the current is given by the curve D/D_0 .

nose resulting from lateral shear. The coefficient (sec^{-1}) represents a coefficient of internal virtual friction which may also absorb bottom friction, such be present [Stommel, 1948; Hansen, 1951]. It is assumed that μ is constant, whereas r is considered constant in the first model and allowed to vary in the cross-stream direction in a second model.

In the following, interest is directed to the horizontal mass transport and the horizontal velocity profile of the Gulf Stream. No attempt is made to derive the vertical velocity profile of the Stream.

Integration of equations (3) over z between (x, y) and $D(x, y)$, followed by cross differentiation and subtraction of the second equation in (3) from the first equation, yields

$$+\frac{\partial p_D}{\partial x} \frac{\partial D}{\partial y} - \frac{\partial p_D}{\partial y} \frac{\partial D}{\partial x} = 0 \quad (4)$$

where $\beta = \partial f / \partial y$ and

$$p_D = g\bar{p}(\zeta - D) \quad (5)$$

and \bar{p} is defined by (2). τ is the wind stress at the sea surface, M the total horizontal mass transport

$$M = (U^2 + V^2)^{1/2}$$

with the components

$$U = \int_D^{\zeta} \rho u \, dz \quad V = \int_D^{\zeta} \rho v \, dz \quad (6)$$

The equation of continuity for the total horizontal mass transport is

$$\partial U / \partial x + \partial V / \partial y = 0 \quad (7)$$

Although the divergence of the flow at individual levels vanishes only under very special conditions, the divergence of the horizontal mass transport in the whole moving layer is zero. The vertical component of a vector-product, $\nabla p_r \times \nabla D$, which appears in the vertically integrated equations (3) after cross differentiation and subtraction, is zero, since the sea surface represents the uppermost isobaric surface for the sea pressure p (constant atmospheric pressure).

From (5) and (2), it follows that (4) can be written as

$$\begin{aligned}
 & V\beta + \text{curl } \tau + r \text{ curl } M \\
 &= gD \left(\frac{\partial \bar{p}}{\partial x} \frac{\partial D}{\partial y} - \frac{\partial \bar{p}}{\partial y} \frac{\partial D}{\partial x} \right) \\
 &+ g\bar{p} \left(\frac{\partial \zeta}{\partial y} \frac{\partial D}{\partial x} - \frac{\partial \zeta}{\partial x} \frac{\partial D}{\partial y} \right) \quad (8)
 \end{aligned}$$

This vorticity-tendency equation contains the vertical components of the vector products, $\nabla \bar{p} \times \nabla D$ and $\nabla \zeta \times \nabla D$. Both of them, and their sum, disappear only under very special conditions which are not necessarily fulfilled in the oceans [Neumann, 1956].

Application to the Gulf Stream. According to the facts and assumptions of the preceding sections, (8) will be applied to a meridional flow that approximates the Gulf Stream with a lower boundary as given by (1).

In this model, essential flow conditions are considered to be constant in the meridional direction. This assumption will eliminate an explanation of why the total mass transport of the Stream from the Straits of Florida northward first increases and later, probably after leaving the region off Cape Hatteras, decreases, as indicated on the mass transport chart shown by *Sverdrup and others* [1946]. Since this inflow of water from the Antilles Current and the outflow of water in the northern regions is not considered an essential part of this model, it is assumed that $\partial U / \partial y$ is zero and the wind stress τ is constant.

With $\bar{p} = \text{constant}$ in the moving Gulf Stream layer and $\partial D / \partial y = 0$, it follows from (8) that

$$r \frac{dV}{dx} + \beta V = g\bar{p} \frac{\partial \zeta}{\partial y} \frac{dD}{dx} \quad (9)$$

where $\partial \zeta / \partial y$ may be considered a frictional slope of the sea surface, which necessarily appears negative in the flow direction.*

* Precise leveling based on 9 tide gage stations along the American east coast indicates a tilting of the mean sea level between Florida and Maine such that the sea surface is about 31 cm higher at Portland, Maine, than at St. Augustine, Florida. It is interesting that this slope is of the same order of magnitude as the 'frictional slope,' $\partial \zeta / \partial y$; however, it has the opposite sign. This raises the question whether the slope in the direction of the Gulf Stream and the slope along the coast can both be explained by frictional effects or whether slight density differences in the flow direction may be of importance.

South of Cape Hatteras, where the Stream flows over the continental shelf, D approximates the bottom topography. North of Cape Hatteras after separation from the coast, the Stream which has adjusted dynamically to this slope as lower boundary, tends to maintain the 'equilibrium' flow pattern and continues with D as a 'free' lower boundary toward the northeast. When the Stream turns toward a more zonal direction a mutual readjustment between the flow and its lower boundary has to be expected. This probably happens shortly after the current leaves the Grand Banks region. The splitting up of the original Stream, the tendency to send a branch to the north along the east coast of Newfoundland, and the broad and diffuse pattern of the North Atlantic Current, as well as the drop of the zero layer toward Greenland according to *Defant* [1941], may be parts of this readjustment of the flow which still remain to be explained. The general results of *Stommel* [1948], *Hidaka* [1949], and *Munk* [1950] on the large scale flow pattern in a whole ocean are of importance in explaining this pattern, as is also the effect of a variable lower boundary on the circulation [*Sverdrup*, 1957]. In the following, however, only the Gulf Stream proper is considered to be approximated by a purely meridional flow.

For a solution of (9), let

$$\epsilon = \frac{x}{h}; \quad A = \frac{g\bar{p}}{r} \frac{\partial \zeta}{\partial y} = \text{constant};$$

Then

$$\gamma = \beta h / r.$$

$$V = e^{-\gamma \epsilon} \left[C + 2A D_0 \int_0^\epsilon \alpha e^{-\alpha^2 + \gamma \alpha} d\alpha \right] \quad (10)$$

With the boundary condition that at $\epsilon = 0$, $V = 0$, the constant C becomes zero, and

$$\begin{aligned}
 V = 2A D_0 e^{-\gamma^2/4} e^{-\gamma \epsilon} & \left[\frac{\gamma}{2} \int_{-\gamma/2}^{\epsilon - \gamma/2} e^{-\eta^2} d\eta \right. \\
 & \left. - \frac{1}{2} (e^{-(\epsilon - \gamma/2)^2} - e^{-\gamma^2/4}) \right] \quad (11)
 \end{aligned}$$

Since V represents the meridional mass transport in the Gulf Stream above the depth D across a section 1 cm by D cm, the average velocity of the vertical columns of water between ζ and D is

$$\bar{v}(x) = V / \bar{p} [H - D_0 \exp(-\epsilon^2)] \quad (12)$$

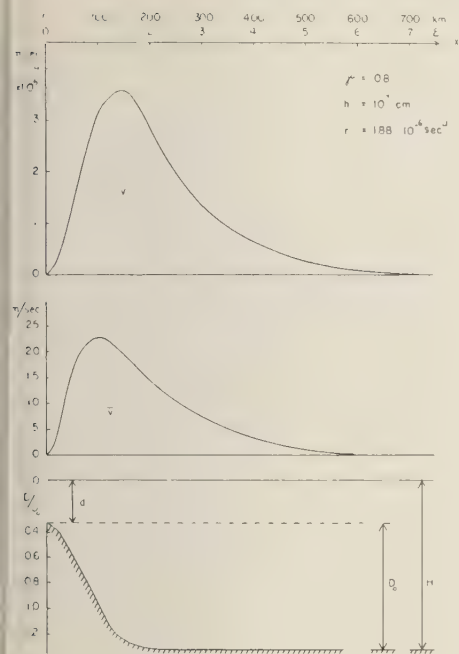


Fig. 5. Profiles of the mass transport V and average velocity \bar{v} across a meridional current computed for $\gamma = 0.8$, $r = 1.88 \times 10^{-6} \text{ sec}^{-1}$. D/D_0 represents the profile of the lower boundary across the current.

With $H = 2000 \text{ m}$, $D_0 = 1500 \text{ m}$, $h = 100 \text{ km}$, $\xi/\partial y \approx 1.5 \times 10^{-7}$ and $r \approx 3.3 \times 10^{-6}$ [Neumann, 1955], the result is that

$$\bar{v}_{\max} = 25.6 \text{ cm/sec}$$

The velocity profile across the Stream according to (12) is presented in Figure 4. It shows some resemblance to the actual profile in the cross-stream direction. A stronger cyclonic shear region near the coast and a less strong anticyclonic shear region toward the Sargasso Sea are indicated. The region of maximum velocity is found at the foot of the continental slope, approximately, or at the foot of the slope of D if D represents a free lower boundary. However, the total width of the current still appears to be too wide.

A narrower velocity profile is obtained if the value of r is assumed to be smaller. With $r = 1.88 \times 10^{-6} \text{ sec}^{-1}$, a horizontal profile for is shown in Figure 5. It is seen that as r

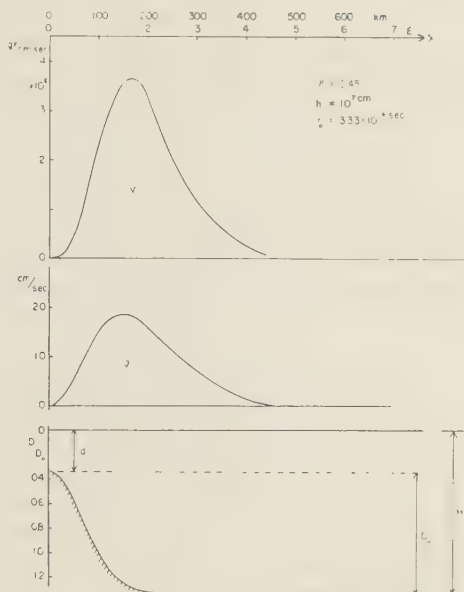


Fig. 6. Profiles of the mass transport V and average velocity \bar{v} across a meridional current computed for $\gamma = 0.45$ and $r = r_0/\epsilon$. The lower boundary of the current, as shown by the curve D/D_0 is the same as in Figure 4.

decreases, \bar{v} also decreases, and there does not seem to be much chance of improving the model by changing r as long as the friction is assumed to be constant across the Stream. However, if a smaller value for the frictional slope is assumed a smaller value for r can be introduced. This keeps the factor $A = (g\bar{\rho}/r)(\partial\xi/\partial y)$ constant, and $\gamma = \beta h/r$ becomes greater, which would result in a narrower profile.

In order to improve the model of a purely meridional flow near a coast, it seems, first of all, that an increase of frictional forces toward the coast would be a reasonable assumption. Not much is known about how the friction should be increased toward the coast. This can be estimated only if more is known about the nature of eddy viscosity in the oceans. As was pointed out by Charney [1955] in connection with the Gulf Stream problem: "An internal friction boundary layer should be introduced to reduce the velocities to zero on the inshore edge of the currents."

A simple assumption about the increase of friction toward the coast can be made, solely to show what the effect of increasing friction on

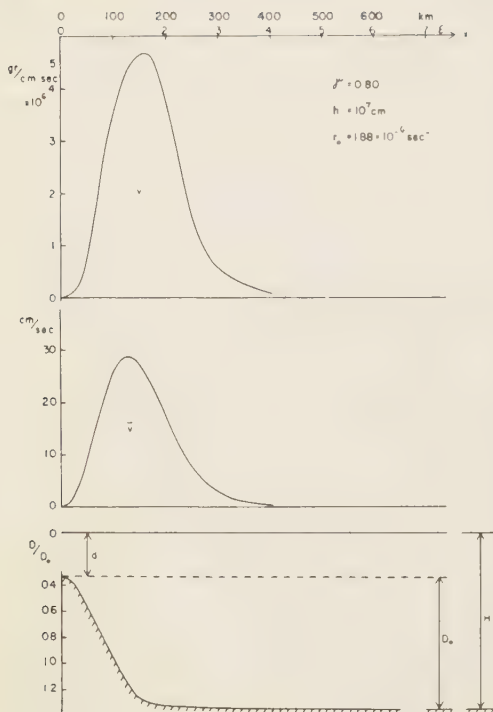


Fig. 7. Profiles of the mass transport V and average velocity \bar{v} across a meridional current computed for $\gamma = 0.80$ and $r = r_0/\epsilon$ with $r_0 = 1.88 \times 10^{-6} \text{ sec}^{-1}$. The curve D/D_0 is the same as in the preceding examples.

the horizontal velocity profile can do. The assumption is given by $r = r_0/\epsilon$, where r_0 is the virtual effective friction coefficient at $\epsilon = 1$, that is, at a distance of 100 km from the coast. In this case it follows from (8) with the same assumptions as in the first model, ($\partial U/\partial y = 0$, $\partial \bar{p}/\partial x = 0$, $\tau = \text{constant}$, and $\partial D/\partial y = 0$), that

$$\frac{dV}{d\epsilon} + \gamma \epsilon V = \frac{A}{r_0} \epsilon^2 \exp(-\epsilon^2) \quad (13)$$

where $\gamma = \beta h/r_0$ and $A = 2g\bar{p}D_0 \partial \xi/\partial y$. The general solution of (13) is

$$V = e^{-\gamma \epsilon^2/2} \left[C + \frac{A}{r_0} \int_0^\epsilon \alpha^2 \cdot \exp - \left(1 - \frac{\gamma}{2} \right) \alpha^2 d\alpha \right] \quad (14)$$

The boundary condition $V = 0$ at $\epsilon = 0$ yields

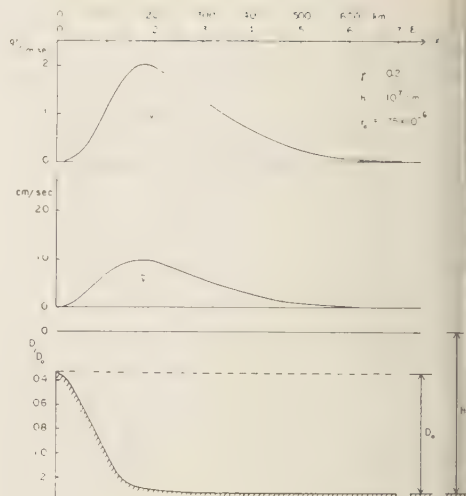


Fig. 8. Profiles of the mass transport V and average velocity \bar{v} across a meridional current computed for $\gamma = 0.2$ and $r = r_0/\epsilon$ with $r_0 = 7.5 \times 10^{-6} \text{ sec}^{-1}$. The curve D/D_0 is the same as in the preceding examples.

$C = 0$, and therefore

$$V = \frac{A}{r_0(2 - \gamma)} \left[\frac{e^{-\gamma \epsilon^2/2}}{(1 - \gamma/2)^{1/2}} \cdot \int_0^{\epsilon(1-\gamma/2)^{1/2}} e^{-\eta^2} d\eta - \epsilon e^{-\epsilon^2} \right] \quad (15)$$

For the same values of the constants as in the evaluation of (12) with (11) as a solution for V , a horizontal profile for the average velocity \bar{v} of the Gulf Stream layer is obtained (Fig. 6). The current appears to be more concentrated toward the coast, and the asymmetry of the profile is lessened. It seems that the increase of friction toward the shallow side of the Stream is more important than the absolute value of the frictional forces. The profile for the horizontal mass transport V is also shown in the figure.

Two more examples are presented in Figures 7 and 8. The dimensions of D are the same as in the preceding cases: that is, $D_0 = 1500 \text{ m}$, $H = 2000 \text{ m}$, and $h = 100 \text{ km}$.

In the case shown in Figure 7, where $\gamma = 0.8$, $r_0 = 1.88 \times 10^{-6}$, the current profile for \bar{v} , as well as the transport profile for V , are more concentrated than in the case shown in Figure 6. The whole current is only 200 to 250 km wide, approximately, and the average peak velocity

the vertical columns of water, \bar{v} , is 29 cm/sec. However, the differences between this profile and the profile shown in Figure 6 are not very significant.

With $\gamma = 0.2$ and $\tau_0 = 7.5 \times 10^{-6}$, the profiles for V and \bar{v} are presented in Figure 8. It is seen that the current broadens markedly, whereas the peak values for the transport V and for \bar{v} are much smaller than in the cases shown in Figures 6 and 7. However, the symmetry of the velocity profile (or transport profile) is clearly indicated in each case.

Acknowledgments. This paper is a report on part of the research carried out under the sponsorship of the Office of Naval Research, Washington, D. C., under Contracts Nonr 285(12) and Nonr 285(03). The author wishes to thank W. J. Pierson, Jr., for helpful discussions and for reading the manuscript of the paper.

REFERENCES

Harney, J. G., The Gulf Stream as an internal boundary layer, *Proc. Natl. Acad. Sci.*, 41, 731-740, 1955.

Defant, A., Die absolute Topographie des physikalischen Meeresniveaus und der Druckflächen, sowie die Wasserbewegungen im Atlantischen Ozean. Dtsche. Atl. Exp. "Meteor," *Wiss. Ergebnisse*, 6, 5, 1941.

Defant, A., Il problema della corrente del Golfo, *Atti del VII Convegno Annuale Roma*, 7-9 Nov. 1957, *Assoc. Geofis. Italiana*, 1957.

Ekman, W., Über Horizontalzirkulation bei wind-erzeugten Meeresströmungen. *Arkiv Math., Astron. Fysik*, 17, 26, 1923.

Ekman, W., Wind-erzeugte Strömungen im Ozean,

Deut. Hydrograph. Z., 4, 161-172, 1951.

Hassan, E. M., On the wind driven ocean circulation, *Deep-Sea Research*, 5, 36-43, 1958.

Hidaka, K., Mass transport in ocean currents and lateral mixing, *J. Marine Research, Sears Foundation*, 8, 1949, (and following papers, *Tokyo Univ. Geophys. Notes*, 1950, 1951).

Iselin, C. O'D., A study of the circulation of the western North Atlantic, *Papers Phys. Oceanog. Meteorol., Mass. Inst. Technol. and Woods Hole Oceanog. Inst.*, 4, 1936.

Martineau, D., The Gulf Stream as an equivalent-barotropic system, Ph.D. thesis (unpublished), Dept. of Meteorol. and Oceanog., New York University, 1958.

Munk, W. H., On the wind driven ocean circulation, *J. Meteorol.*, 7, 79-93, 1950.

Neumann, G., On the dynamics of wind driven ocean currents, *Meteorol. Papers*, 2, 33 pp., 1955.

Neumann, G., Zum Problem der dynamischen Bezugsfläche, insbesondere im Golfstromgebiet. *Deut. Hydrograph. Z.*, 9, 66-78, 1956.

Stommel, H., The westward intensification of wind-driven ocean currents, *Trans. Am. Geophys. Union*, 29, 202-206, 1948.

Sverdrup, H. U., M. W. Johnson, and R. W. Fleming, *The Oceans*, Prentice-Hall, New York, 1946.

Sverdrup, H. U., Oceanography, in *Handbuch der Physik*, Band XLVIII, Geophysik II, pp. 658, 664-666. Herausgeg. von S. Flügge, Springer Verlag, Berlin-Göttingen-Heidelberg, 1957.

Swallow, J. C., and L. W. Worthington, Measurements of deep currents in the western North Atlantic, *Nature*, 179, 1183-1184, 1957.

von Arx, W., Notes on the surface velocity profile and horizontal shear across the width of the Gulf Stream, *Tellus*, 3, 211-214, 1952.

(Manuscript received July 6, 1959.)

On the Statistical Prediction of Ocean Temperatures

GUNNAR I. RODEN AND GORDON W. GROVES

*Scripps Institution of Oceanography
La Jolla, California*

Abstracts. Past records of sea-surface temperature are examined and related to wind and solar radiation. There is practically no relation between radiation and sea-surface temperature, but there is a noticeable but small relation between wind and temperature. The best estimate of future values of temperature can be made from past values of temperature, and the use of past wind values slightly improves the prediction. In any case, optimum prediction from statistical considerations alone (Wiener's technique) can account for only about 40 per cent of the variance in the next month's average temperature.

Introduction. The prediction of sea-surface temperature a few months into the future is of interest to both meteorologists and oceanographers and has many practical applications, including fisheries forecasts. In the past, several methods have been used to estimate future temperatures; these are summarized by Wadsworth [1951], Namias [1953], and Takenouti [1957]. In the present investigation we deal with the prediction of temperature from past records of wind, temperature, and radiation, using Wiener's [1949] method.

In statistical prediction no attempt is made to use dynamical relations, and an exact value of temperature at any future time cannot be furnished. However, an estimate of the most probable value and the limits within which it is likely to fall can be given. Here it is desired to evaluate the degree of success attainable by the purely statistical approach. Any method based on dynamical considerations which does not give a better estimate than can be obtained by statistical methods (that is, by chance) will not have contained the essential mechanism. Actually, as is shown below, the accuracy of purely statistical prediction of sea-surface temperature is only fair. A similar analysis was carried out by Pierson [1958] in the North Atlantic with somewhat better results as to the accuracy of prediction.

Two ocean areas have been selected for testing the method. The first is a 5° square, 45°N to

50°N , 125°W to 130°W , just off the coast of Washington; the other, 25°N to 30°N , 140°W to 145°W , is in the NE trade wind region about halfway between San Francisco and Honolulu. Eighteen-year records of mean monthly sea-surface temperature, cloudiness [*Imperial Marine Observatory, Kobe*, 1921–1938], and sea-level atmospheric pressure [*U. S. Weather Bureau* 1899–1958] were used. The cloudiness was converted into incoming radiation by means of Budyko's [1956] tables, and pressure was converted into wind by assuming geostrophic equilibrium. To eliminate seasonal variations, differences from the long-term monthly mean were taken (that is, from any given value the average for the same month in all years was subtracted). The resulting anomalies are shown in Figure 1. Several features are evident in the records: (1) they are somewhat similar in general character; (2) the temperature anomalies do not change their sign as often as do those of wind and radiation; (3) the records appear to be stationary. One can expect that the autocorrelation of temperature will show a slower decay than will that of radiation and wind.

It must be emphasized that taking monthly averages of daily values presupposes a smoothing process. This smoothing generally increases the persistence over that of the original unsmoothed records [Munk, 1959].

Analysis of the data. Correlation functions and spectra of the time series shown in Figure 1 were computed on an IBM 704 electronic computer, using the Tukey method [Tukey, 1949; Blackman and Tukey, 1958]. The analysis

Contribution from the Scripps Institution of Oceanography. New Series.

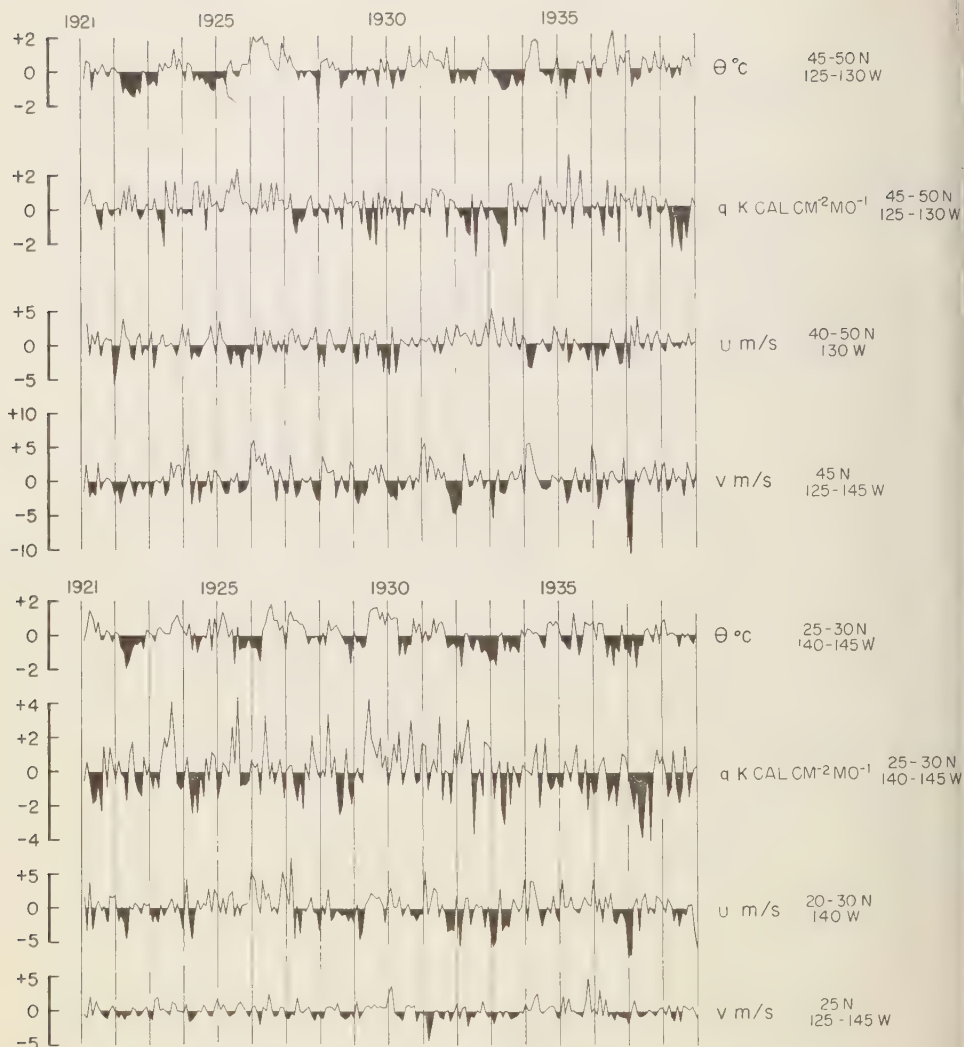


Fig. 1. Time series of anomalies of temperature (θ), east (u), and north (v) components of wind, and radiation (q) for the region off the coast of Washington (top) and the NE trade wind region (bottom).

was carried out for $n = 216$ observations and $m = 24$ lags or spectral estimates. The number of degrees of freedom,

$$\nu = 2n/m = 18$$

determines the statistical reliability of the spectral estimates. The definitions of the various

functions, autocorrelations, cross correlations, energy spectra, coherence, and phase are given by Munk, Snodgrass, and Tucker (in press).

The normalized autocorrelations of temperature, radiation, and east and north wind components and radiation, denoted by $\phi_{\theta\theta}$, ϕ_{uu} , ϕ_{vv} , ϕ_{qq} , are shown in Figures 2a and 3a.

The statistical reliability of the autocorrela-

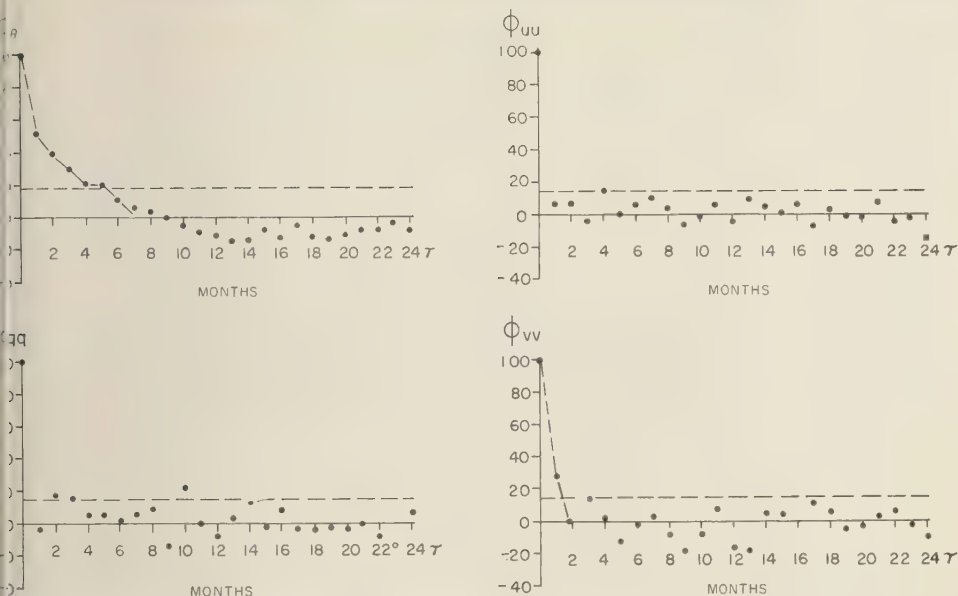


Fig. 2a. Normalized autocorrelation of temperature, wind components, and radiation anomalies in the region off the coast of Washington.

ons can be estimated by the inequality

$$\sigma(\tau)^2 \leq \sigma_m^2 = \frac{4}{T} \int_0^\infty \phi^2(\tau') d\tau'$$

which holds for all τ , where σ^2 is the (normalized) variance in the autocorrelation, determined from random finite samples of an infinite time series [Laning and Battin, 1956, p. 163], and τ' is the record length. The quantity σ_m is indicated by a dashed line on the autocorrelation plots in order to give a rough idea of the significance. It is seen that the autocorrelation of temperature differs significantly from zero for lags up to 5 or 6 months, indicating a rather high persistence of temperature. The autocorrelation of wind, on the other hand, does not differ significantly from zero for lags greater than 1 month. The autocorrelation of radiation as determined from cloudiness is not very persistent, as could be expected.

The cross correlations between pairs of variables are shown in Figures 2b and 3b. Significance levels for the cross correlations have not yet been determined, but experience suggests that the erratic variations at large (positive or negative) lag times are not significant. There-

fore, we might reason that any peak rising appreciably above the level of those variations might be significant. Such correlations occur only for lag times near zero. Wind and temperature are the best correlated variables; there is hardly a significant relation between any other pair. It should be noted that wind leads temperature for negative lag time. The dotted curves in the correlation graphs show the smoothed correlations that were used in the prediction calculations. The smoothing was carried out by a simple moving-weighted-average, $0.25 \phi_{-1} + 0.50 \phi_0 + 0.25 \phi_1$, applied at each lag in the significant part of the correlation. Outside the significant part the correlations were set equal to zero.

The energy spectra, coherences, and phase relationships are shown in Figures 4 and 5. The energy spectra of temperature show that most of the energy is concentrated in the low-frequency part of the spectrum, between zero and 2 cycles per year (cpy). This is consistent with the slow decay of the temperature autocorrelation; both indicate that future temperatures can be forecast from past temperatures. There are no obvious periodicities in the spectrum

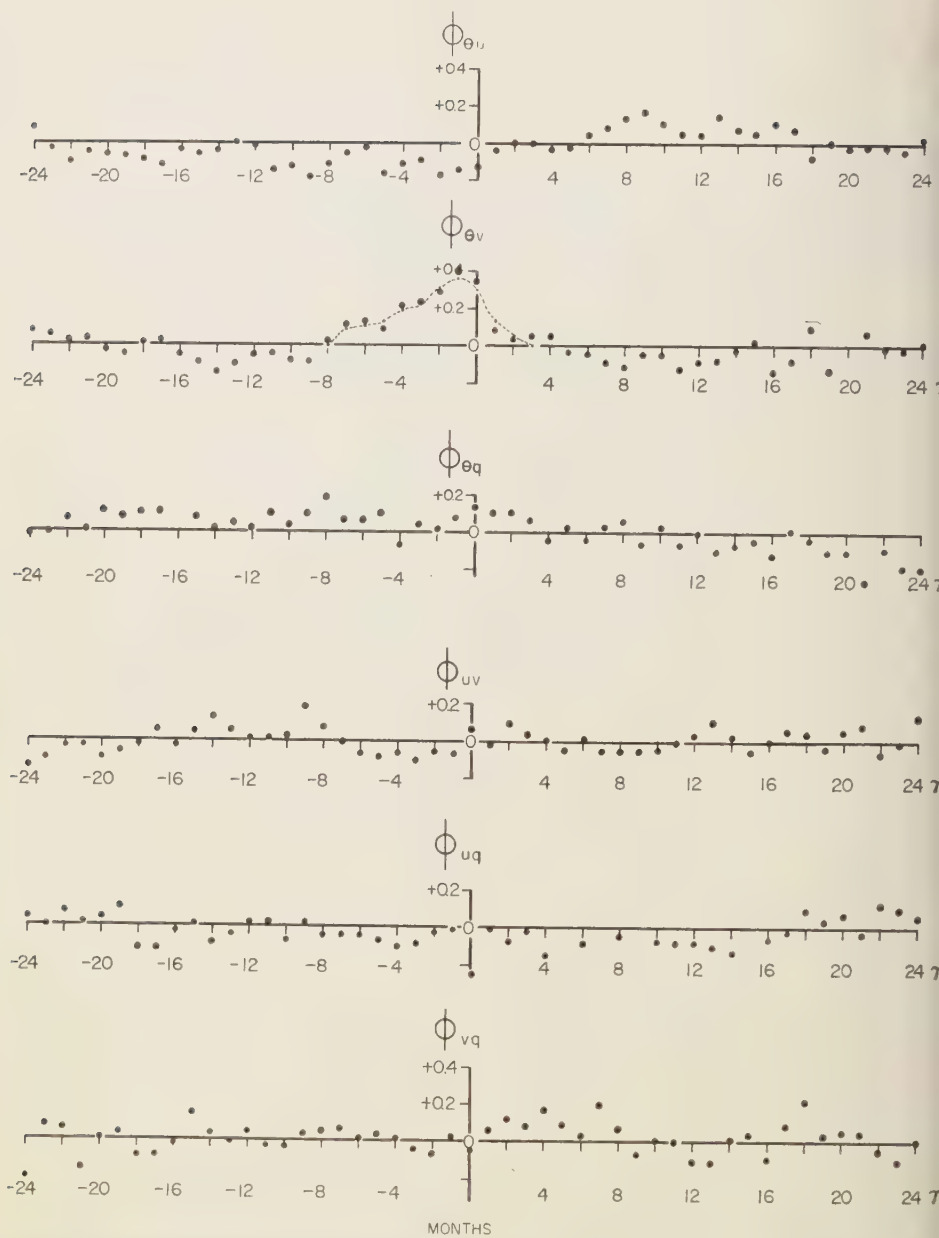
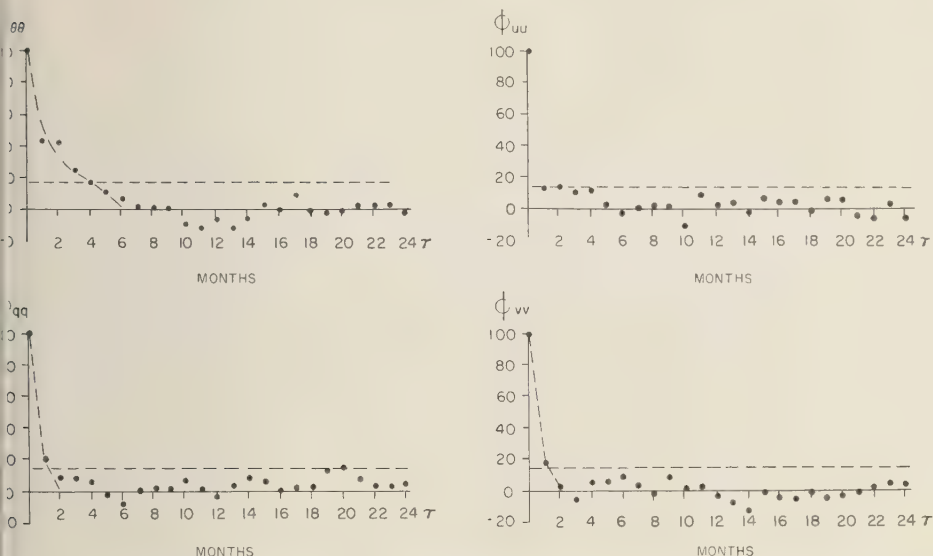


Fig. 2b. Normalized cross correlations of temperature, wind components, and radiation anomalies in the region off the coast of Washington.



g. 3a. Normalized autocorrelations of temperature, wind components, and radiation anomalies in the NE trade wind region.

cept for a broad-band maximum in the vicinity of 0.25 cpy. The low energy level in the higher-frequency part of the spectrum and the fact that the input data consist of monthly averaged data would indicate that aliasing is not a serious problem.

The spectra of wind are almost constant throughout the frequency range or show only a slight decrease of energy with increasing frequency. In other words, the wind record deviates significantly from white noise. This is consistent with the rapid decay of the wind autocorrelation and indicates that prediction of wind from past wind records would yield very poor results. No important periodicities are indicated in the wind spectra except for dubious peaks around 3.0 and 5.5 cpy. The spectrum of radiation appears to be even more white than the wind. The 95 per cent significance level of all the energy spectra lies between 0.57 and 2.20 times the observed value [Blackman and Tukey, 1958].

The 95 per cent significance levels for the coherence are indicated by the dashed lines in Figures 4b and 5b. It is seen that the coherence between temperature and wind is significant in the low-frequency part of the spectrum (zero

to 1 cpy). There appears to be significant coherence again in medium frequencies, but this may be doubtful. Perhaps an even better indication of significant coherence between records is given by the continuity in the phase relation as a function of frequency. If the phase values vary considerably, it probably means that the records are not very coherent. It is noted that the only part of the spectrum showing a persistent phase relation between temperature and wind is for low frequencies, for which both are approximately in phase. In the NE trade wind region easterly and southerly winds are associated with high temperatures, for low frequencies. In the Washington region easterly winds are associated with low temperatures, as is indicated by phase difference of about 180° in the low-frequency range, and the southerly wind accompanies higher temperatures.

The coherence between temperature and radiation is generally very poor, although there does appear to be some coherence in the NE trade wind region for very low frequencies, between zero and 0.5 cpy.

Prediction. We shall consider the methods outlined by Wiener [1949], taking each time series to be a stationary ergodic process. This

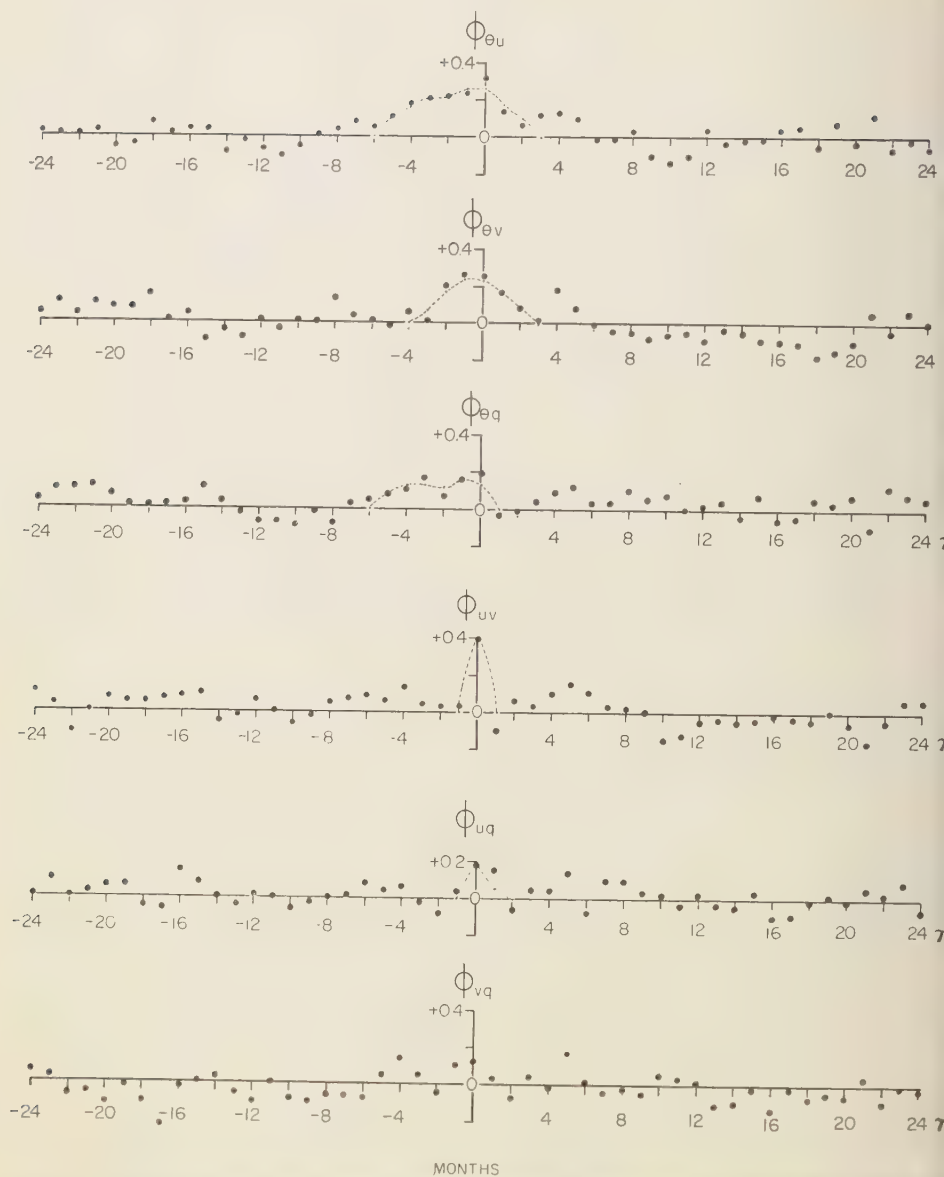


Fig. 3b. Normalized cross correlations of temperature, wind components, and radiation anomalies with the NE trade wind region.

allows us to equate ensemble and time averages and assures us that the correlation functions depend only on the time displacement (τ), and not on any particular origin in time. The pre-

dicted temperature anomaly θ^* at any future month is taken to be a linear combination of the monthly anomalies of all variables for the present and a few past months. Accordingly,

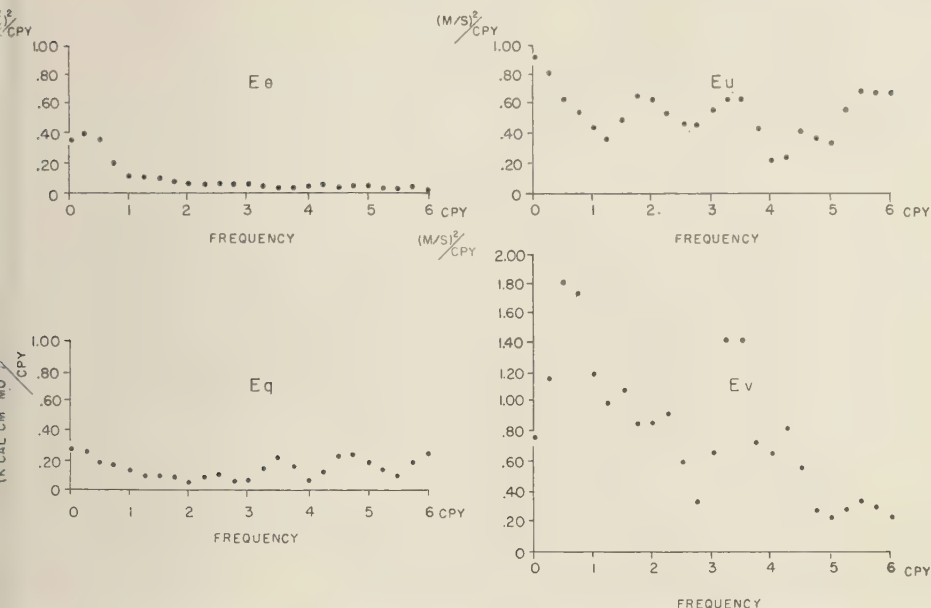


Fig. 4a. Energy spectral density E of temperature, wind, and radiation anomalies, Washington region.

$$\begin{aligned} \epsilon(t + \alpha) = & \sum_{i=0}^n [a_i \theta(t - i) \\ & + b_i \lambda_{\theta u} u(t - i) + c_i \lambda_{\theta v} v(t - i) \\ & + d_i \lambda_{\theta q} q(t - i)] \end{aligned} \quad (1)$$

where t is the time measured in months; θ , u , v , and q are the observed anomalies of temperature, two wind components, and radiation, respectively; α is the number of months ahead for which the prediction is to be made; and n is the number of past months taken into account. The quantities

$$\lambda_{\theta u} = [\langle \theta^2 \rangle / \langle u^2 \rangle]^{1/2}, \text{ etc.} \quad (2)$$

are introduced to keep the coefficients a_i , b_i , c_i , and d_i dimensionless. The symbol $\langle \rangle$ indicates a long time average, which is approximated, in the present case, by averaging over the available values.

The coefficients are to be determined in such manner that the mean-square error is minimized. The quantity

$$\epsilon^2 = \langle (\theta^* - \theta)^2 \rangle / \langle \theta^2 \rangle \quad (3)$$

is a useful dimensionless measure of the pre-

diction error. Note that $0 \leq \epsilon^2 \leq 1$; if $\epsilon^2 = 0$ the prediction is perfect, whereas if $\epsilon^2 = 1$ the prediction scheme does no better than if it always predicted zero anomalies. Combining (1), (2), and (3), ϵ^2 can be expressed in matrix notation,

$$\epsilon^2 = 1 - 2X'N + X'MX \quad (4)$$

where X is the unknown column vector.

$$\begin{bmatrix} a_1 \\ \vdots \\ b_1 \\ \vdots \\ c_1 \\ \vdots \\ d_1 \end{bmatrix}$$

N is the column vector

$$\begin{bmatrix} \phi_{\theta\theta}(-\alpha - i) \\ \vdots \\ \phi_{\theta u}(-\alpha - i) \\ \vdots \\ \phi_{\theta v}(-\alpha - i) \\ \vdots \\ \phi_{\theta q}(-\alpha - i) \end{bmatrix}$$

and M is the matrix

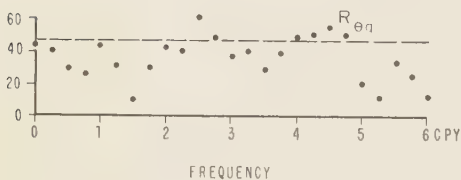
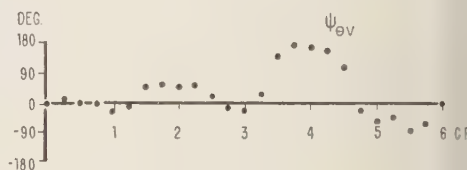
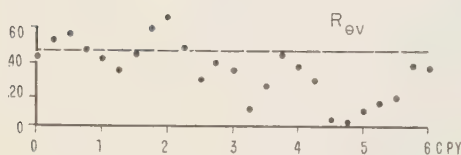
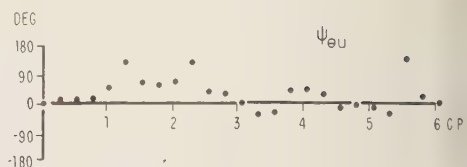
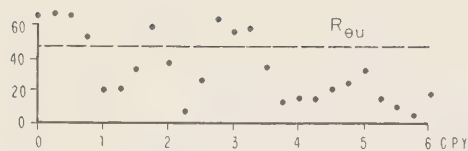


Fig. 4b. Coherence and phase relationships off the coast of Washington.

$$\begin{bmatrix} \phi_{\theta\theta}(i-k) & \phi_{u\theta}(i-k) & \phi_{v\theta}(i-k) & \phi_{n\theta}(i-k) \\ \phi_{\theta u}(i-k) & \phi_{uu}(i-k) & \phi_{vu}(i-k) & \phi_{nu}(i-k) \\ \phi_{\theta v}(i-k) & \phi_{uv}(i-k) & \phi_{vv}(i-k) & \phi_{nv}(i-k) \\ \phi_{\theta n}(i-k) & \phi_{un}(i-k) & \phi_{vn}(i-k) & \phi_{nn}(i-k) \end{bmatrix}$$

in which $i, k = 0, 1, \dots, n$;

and X' is the transpose of X . The ϕ are normalized correlation functions, defined by

$$\phi_{\theta u}(\tau) = \frac{\langle \theta(t)u(t+\tau) \rangle}{\sqrt{\langle \theta^2 \rangle \langle u^2 \rangle}} \quad (5)$$

For $\tau > 0$ the primary record is related to future values of the secondary record according to (5).

To minimize the mean-square error, expression (4) is differentiated with respect to each unknown coefficient and the result set equal to zero. This may be loosely considered as differentiating with respect to the vector X (like taking the gradient) and equating to zero [Frazer

and others, 1938, p. 48]. We obtain

$$MX = N \quad (6)$$

representing a system of linear algebraic equations which can be solved to yield values of the coefficient for optimum prediction. From the solution

$$X = M^{-1}N$$

we can calculate the minimized error by substitution into (4),

$$\epsilon_m^2 = 1 - X'N = 1 - N'M^{-1}N \quad (7)$$

making use of the fact that M is a symmetrical matrix. It is to be noted that the error and the coefficients depend on the number of past and present values, $n+1$, upon which the prediction is based, as well as on the number of months ahead, α , for which the prediction is to be made.

Suppose that we are to make a prediction of temperature from its own present and past

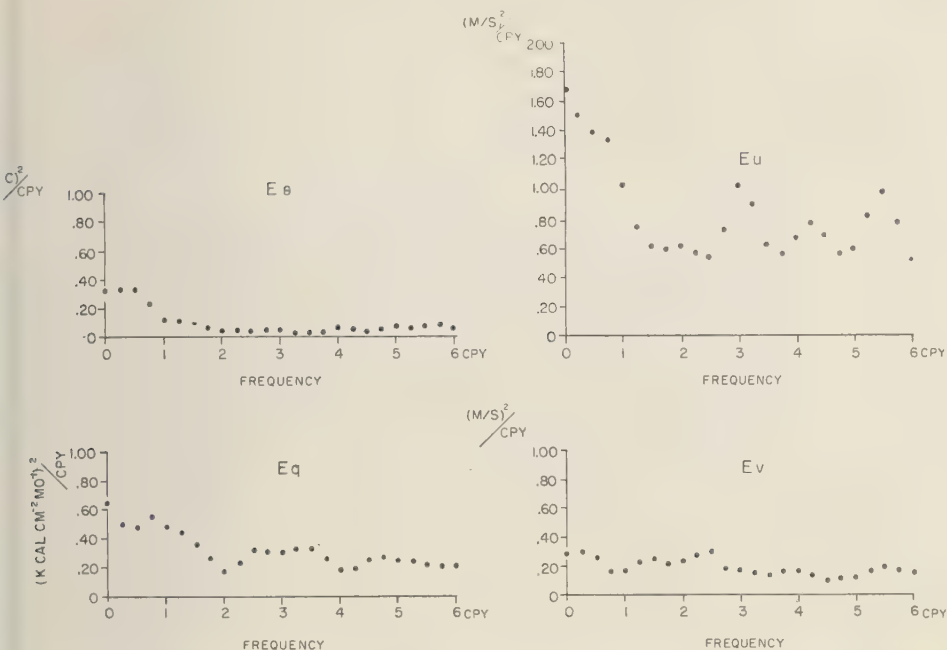


Fig. 5a. Energy spectral density E of temperature, wind, and radiation anomalies, NE trade wind region.

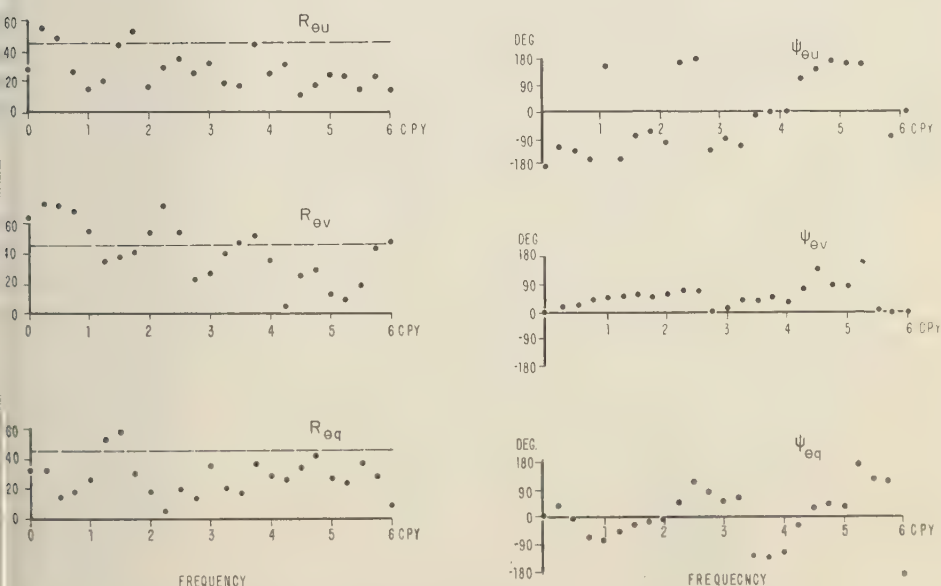


Fig. 5b. Coherence and phase relationships in the NE trade wind region.

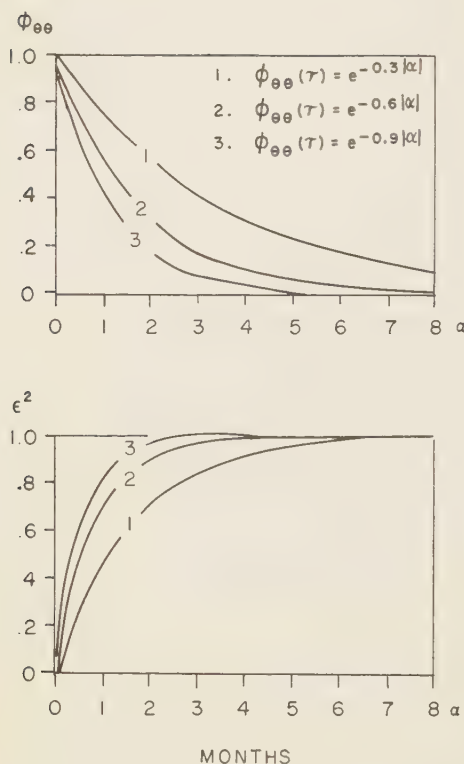


Fig. 6. Autocorrelations and mean-square error for three examples of decay rates in a Markov process.

and suppose also that the autocorrelation of temperature is adequately expressed by

$$\phi_{\theta\theta}(\tau) = e^{-\lambda|\tau|}, \lambda > 0$$

It is evident from Figures 2a and 3a that the autocorrelations do not differ significantly from the above expression with appropriate decay rates λ . Time series having such exponential autocorrelations are known as Markov processes, and it is well known that taking past values into account in the prediction scheme does not improve the prediction. In fact, the prediction coefficients turn out to be

$$a_0 = e^{-\lambda\alpha} \quad a_1 = a_2 = \dots = 0$$

and

$$\epsilon_m^2 = 1 - e^{-2\lambda\alpha}$$

Three examples of exponential autocorrelation and the corresponding mean-square errors are shown in Figure 6. The persistence of most temperature records is within the range of curves 1 and 3; therefore, successful prediction cannot be expected far into the future. If the prediction is based on other variables also, even though the autocorrelations and cross correlations are of the simple exponential type, inclusion of past values improves the prediction unless all the decay rates are equal. One must therefore conclude that prediction from temperature alone might as well be based on the present value, whereas prediction from temperature and other variables will be improved somewhat by taking account of some past values of all variables.

Results. The prediction coefficients and the mean-square error were determined for a number of cases. It was desired to find out how the coefficients and the mean-square error depend on α and n and how the number of variables considered in the prediction schemes affects the outcome of prediction. The numerical results of this investigation are given in Tables 1 and 2, and certain important features are illustrated in Figures 7 and 8. Three characteristic aspects stand out immediately: (1) the mean-square error increases rapidly with increasing future month α , (2) the mean-square error rapidly approaches a limiting value with the number of past months taken into account, and (3) the mean-square error is hardly affected by the inclusion of radiation in the prediction scheme. The best prediction can be made by considering both temperature and wind. If only one or the other is considered, temperature can be predicted better from past temperatures than from past winds. There is some improvement in prediction if account is taken of the previous month in addition to the present one, but not much further refinement is gained by including more past data. Useful prediction seems almost to be limited to 1 to 2 months into the future. The smallest mean-square error in our investigation is 0.63 (compared with about 0.5 achieved by Pierson [1958] for the North Atlantic); this is 37 per cent of the temperature variance can be predicted 1 month ahead. The correlation coefficient $R = \sqrt{1 - \epsilon^2}$ between the observed and predicted month is hence 0.61. The smallest mean-square error for 2 months ahead is

TABLE 1. Prediction Coefficients and Mean-Square Error for Washington Region*

Coeff.	α					
	1	2	3	4	5	6
	$\theta, v, 4$					
a_0	0.32	0.21	0.20	0.11	0.18	0.12
a_1	0.10	0.10	0.01	0.13	0.05	-0.07
a_2	0.08	-0.01	0.09	0.03	-0.10	-0.02
a_3	0.01	0.09	-0.02	-0.12	-0.03	0.00
c_0	0.22	0.22	0.14	0.14	0.07	0.07
c_1	0.09	0.03	0.06	0.01	0.01	0.06
c_2	0.02	0.07	0.01	0.02	0.05	-0.02
c_3	0.05	0.01	0.03	0.04	-0.03	0.01
ϵ^2	0.65	0.78	0.87	0.91	0.94	0.97

Coeff.	α				Coeff.	α			
	1	2	3	4		1	2	3	4
	$\theta, 4$					$v, 4$			
a_0	0.42	0.31	0.26	0.16	c_0	0.31	0.28	0.19	0.17
a_1	0.13	0.13	0.04	0.15	c_1	0.19	0.10	0.12	0.06
a_2	0.08	-0.01	0.10	0.04	c_2	0.12	0.14	0.07	0.07
a_3	-0.02	0.06	-0.04	-0.13	c_3	0.16	0.09	0.09	0.07
ϵ^2	0.71	0.83	0.89	0.93	ϵ^2	0.77	0.85	0.92	0.95
	$\theta, 2$					$v, 2$			
a_0	0.44	0.32	0.27	0.15	c_0	0.29	0.27	0.18	0.16
a_1	0.16	0.15	0.07	0.12	c_1	0.23	0.14	0.14	0.09
ϵ^2	0.71	0.83	0.91	0.95	ϵ^2	0.83	0.89	0.91	0.96
	$\theta, v, 2$					$\theta, v, 1$			
a_0	0.35	0.30	0.22	0.10	a_0	0.45	0.33	0.26	0.17
a_1	0.16	0.19	0.06	0.12	c_0	0.22	0.21	0.14	0.14
c_0	0.21	0.26	0.13	0.14	ϵ^2	0.69	0.80	0.89	0.93
c_1	0.08	-0.24	0.06	0.02					
ϵ^2	0.66	0.80	0.88	0.93					
	$\theta, 1$					$v, 1$			
a_0	0.52	0.39	0.31	0.21	c_0	0.36	0.31	0.22	0.19
ϵ^2	0.73	0.85	0.90	0.96	ϵ^2	0.87	0.90	0.95	0.96

Temperature θ ; wind u, v ; radiation q ; $\lambda_{\theta} = 0.33^\circ\text{C}/\text{m sec}^{-1}$. The number after the variable refers to the number of past months taken into account; α refers to future time of prediction (in months).

6; that is, only 24 per cent of the temperature variance can be accounted for. For 3 months ahead the corresponding values are 0.86 and 14 per cent.

The prediction coefficients a, b, c , and d do not show such simple behavior as does the mean-square error. The temperature coefficients a increases slowly with increasing α , and very rapidly as increasing numbers of past months

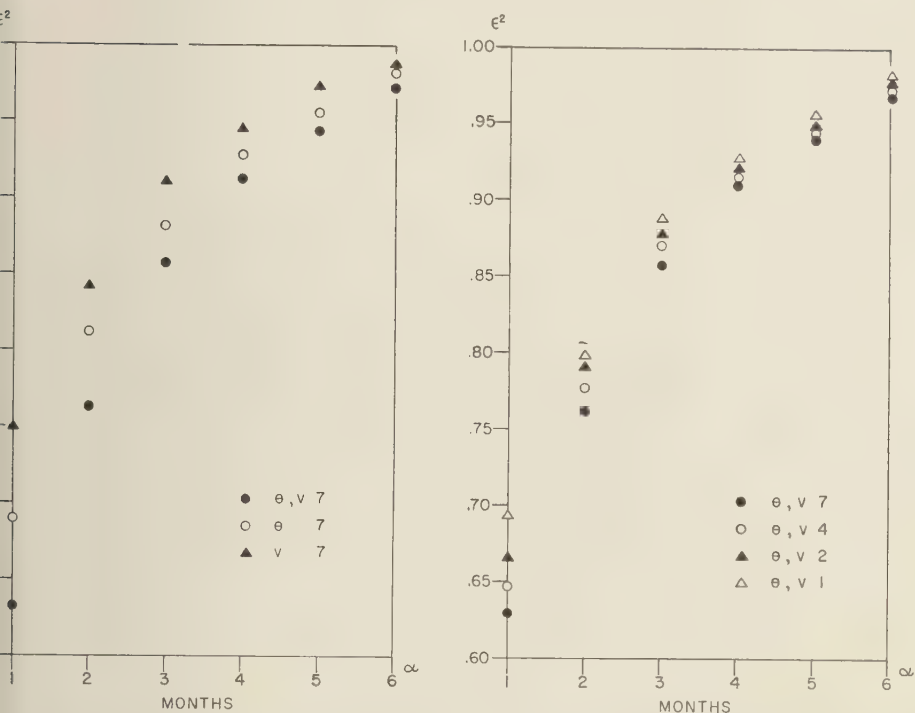
are taken into account. This suggests that future values of temperature depend primarily on the present value and very little on the past, a result which is characteristic of a Markov process. The prediction coefficients of wind decrease slowly as the number of past months is increased, but their dependence on α is erratic. The coefficients of radiation are negligibly small.

In order to investigate how well the predic-

TABLE 2. Prediction Coefficients and Mean-Square Error for NE Trade Wind Region*

Coeff.	α					Coeff.	α			
	1	2	3	4	5		1	2	3	4
$\theta, u, v, q, 4$						$\theta, u, v, q, 2$				
a_0	0.37	0.16	0.16	0.14	0.13	a_0	0.40	0.19	0.15	0.12
a_1	0.00	0.04	0.04	0.05	-0.06	a_1	0.04	0.07	0.04	0.01
a_2	0.03	0.02	0.04	-0.07	-0.01	b_0	0.09	0.11	0.16	0.16
a_3	0.03	0.00	-0.07	-0.01	0.00	b_1	0.08	0.15	0.14	0.08
b_0	0.11	0.13	0.16	0.16	0.10	c_0	0.14	0.10	0.00	-0.07
b_1	0.11	0.16	0.14	0.08	-0.02	c_1	0.01	-0.07	-0.10	-0.05
b_2	0.13	0.14	0.07	-0.03	-0.01	d_0	0.05	0.04	0.05	0.06
b_3	0.09	0.06	-0.04	-0.02	-0.01	d_1	0.09	0.09	0.08	0.05
c_0	0.11	0.08	-0.02	-0.09	-0.06	ϵ^2	0.69	0.84	0.88	0.93
c_1	0.03	-0.04	-0.09	-0.05	0.00	$\theta, u, v, 2$				
c_2	-0.05	-0.08	-0.05	0.00	-0.01	a_0	0.44	0.23	0.17	0.13
c_3	-0.07	-0.03	0.01	0.01	0.01	a_1	0.05	0.08	0.05	0.00
d_0	0.07	0.05	0.05	0.06	0.05	b_0	0.12	0.16	0.20	0.17
d_1	0.01	0.03	0.04	0.03	-0.02	b_1	0.13	0.18	0.15	0.10
d_2	0.04	0.05	0.04	-0.01	0.00	c_0	0.03	-0.03	-0.11	-0.09
d_3	0.04	0.03	-0.02	-0.01	-0.01	c_1	-0.11	-0.14	-0.09	-0.06
ϵ^2	0.65	0.81	0.87	0.93	0.97	ϵ^2	0.71	0.85	0.89	0.94
$\theta, u, v, 4$						$\theta, u, v, q, 1$				
a_0	0.38	0.18	0.17	0.15	0.13	a_0	0.46	0.26	0.19	0.14
a_1	-0.01	0.04	0.04	0.05	-0.06	b_0	0.08	0.11	0.16	0.15
a_2	0.03	0.02	0.04	-0.07	-0.01	c_0	0.10	0.07	-0.04	-0.09
a_3	0.03	0.00	-0.08	-0.01	0.00	d_0	0.14	0.10	0.10	0.09
b_0	0.12	0.13	0.17	0.17	0.11	ϵ^2	0.70	0.87	0.91	0.94
b_1	0.11	0.17	0.15	0.09	-0.02	$\theta, u, v, 1$				
b_2	0.13	0.15	0.09	-0.02	-0.01	a_0	0.48	0.27	0.19	0.14
b_3	0.10	0.07	-0.04	-0.02	-0.01	b_0	0.10	0.12	0.18	0.17
c_0	0.10	0.08	-0.03	-0.09	-0.06	c_0	0.08	0.07	-0.04	-0.11
c_1	0.02	-0.05	-0.10	-0.05	0.00	ϵ^2	0.70	0.87	0.91	0.95
c_2	-0.05	-0.09	-0.05	0.00	0.00	$\theta, 2$				
c_3	-0.07	-0.04	0.02	0.01	0.01	a_0	0.48	0.27	0.21	0.17
ϵ^2	0.67	0.82	0.88	0.93	0.98		0.07	0.10	0.07	0.01
$u, v, 4$						a_1	0.73	0.89	0.94	0.97
b_0	0.20	0.18	0.21	0.20	0.13	ϵ^2	$\theta, 1$			
b_1	0.19	0.22	0.20	0.12	0.00	a_0	0.52	0.32	0.24	0.17
b_2	0.22	0.20	0.12	0.00	0.00	ϵ^2	0.73	0.90	0.94	0.97
b_3	0.20	0.12	-0.03	0.00	0.00	$\theta, 4$				
c_0	0.15	0.10	0.00	-0.07	-0.05	a_0	0.52	0.36	0.29	0.23
c_1	0.07	-0.02	-0.07	-0.04	0.01	a_1	-0.07	-0.15	-0.12	-0.11
c_2	-0.01	-0.07	-0.04	0.01	0.00	a_2	0.10	0.14	0.12	-0.10
c_3	-0.08	-0.04	0.01	0.00	0.00	a_3	0.01	-0.01	-0.09	-0.02
ϵ^2	0.78	0.85	0.91	0.95	0.99	ϵ	0.73	0.90	0.94	0.97

* Temperature θ ; wind u, v ; radiation q . $\lambda_{\theta u} = 0.35^\circ\text{C}/\text{m sec}^{-1}$; $\lambda_{\theta v} = 0.74^\circ\text{C}/\text{m sec}^{-1}$; $\lambda_{\theta q} = 0.5^\circ\text{C cm}^{-2} \text{ mo}^{-1}$. The number after the variables refers to the number of past months taken into account. α refers to future time of prediction (in months).



7. The mean-square error and its dependence on the number of variables used, number of past months considered, and future time of prediction α for the Washington region.

works on independent data, use was made of the temperature records at Amphitrite Point, C. (48°55'N, 125°32'W), and north-south wind index along 45°N, 125° to 145°W during the period 1949-1958. The temperature (°C) the following month was predicted from the temperature and wind (m sec^{-1}) of the present month by the formula

$$\theta^*(1) = 0.45\theta(0) + 0.22\lambda_{\theta} V(0)$$

where $\lambda_{\theta} = 0.33^\circ\text{C/m sec}^{-1}$. It is seen that the observed and the predicted time series (Fig. 9) agree well with respect to the low-frequency part, but that the high-frequency part cannot be predicted very well. Also, the predicted temperature has a smaller rms deviation from the observed than from the observed, which is a general characteristic of linear regression methods. The computed error is $\epsilon^2 = 0.59$, indicating that 59 per cent of the variance has been accounted

for. This is better than the theoretical capability of the prediction, and it must be assumed that an error correspondingly larger than the theoretical value would have been just as likely.

Acknowledgments. The authors benefited considerably from many fruitful discussions with Walter Munk. Maurice Blackburn, John Isaacs, and Joseph L. Reid, Jr., read the manuscript and offered many valuable suggestions. The authors are also indebted to Gaylord Miller and to Convair Astronautics of San Diego, especially to Mr. Bruno Witte of that establishment, who allowed us to use their machine programs and who performed many of the numerical computations on their 704 electronic computer. The remaining desk computations were painstakingly done by Joan Connell.

This work was supported in part by the Scripps Tuna Oceanography Research and the Marine Life Research programs. Funds for the former project were provided by the U.S. Bureau of Commercial Fisheries, Department of the Interior, Contract 14-19-008-9354.

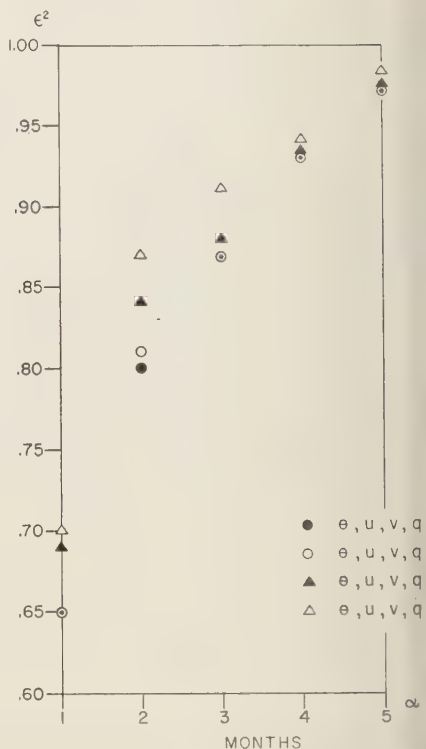
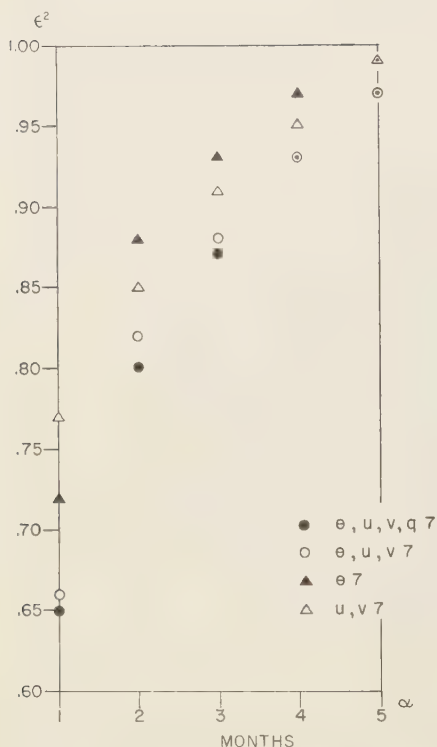


Fig. 8. The mean-square error and its dependence upon the number of variables used, number of months considered, and future time of prediction α for the NE trade wind region.

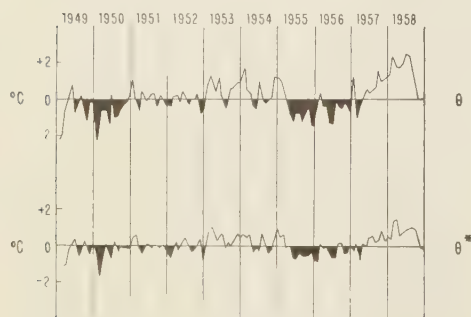


Fig. 9. Observed and predicted temperatures for Amphitrite Point, B. C.

REFERENCES

- Blackman, R. B., and J. W. Tukey, The measurement of power spectra from the point of view of communications engineering, *Bell System Tech. J.* 37, Part I, 185-282, Part II, 485-570, 1958.
- Budyko, M. I., *Teplovoi Balans Zemnoi Poverkhnosti*, GIMIZ, Leningrad, 256 pp., 1956.
- Frazer, R. A., J. A. Duncan, and A. R. Collingwood, *Elementary Matrices*, Cambridge Univ. Press, 416 pp., 1938.
- Imperial Marine Observatory, Kobe, Japan, *Tables of mean atmospheric pressure, cloudiness, air temperature and sea surface temperature of the North Pacific Ocean and the neighboring seas*, Annual publication of monthly values, 1925-1938.
- Laning, J. H., and R. H. Battin, *Random Processes in Automatic Control*, McGraw-Hill Book Company, New York, 434 pp., 1956.
- Munk, W. H., *Smoothing and persistence*, *Meteorol.*, in press.
- Munk, W. H., F. E. Snodgrass, and M. J. Tucker, *Spectra of Low Frequency Ocean Waves*, Univ. of Calif. Press, Los Angeles, 1959.
- Namias, J., Thirty-day weather forecasting: a review of a ten-year experiment, *Meteorological Monographs*, 2(6), Am. Meteorol. Soc., Boston, Mass., 83 pp., 1953.
- Pierson, W. J., *The use of time series concepts and spectral analysis in the study of long range*

forecasting problems, N. Y. Univ., Dept. Meteorol. and Oceanog., 78-100, 1958.

Kenouti, Y., Long range oceanographic forecast, historical review and future problem, *Oceanog. Mag., Japan*, 9, 1-12, 1957.

Key, J. W., The sampling theory of power spectrum estimates, *Symp. Appl. Autocorrelation Analysis to Phys. Probl.*, NAVEXOS-P-735, Office of Naval Research, Washington, D. C., 1949.

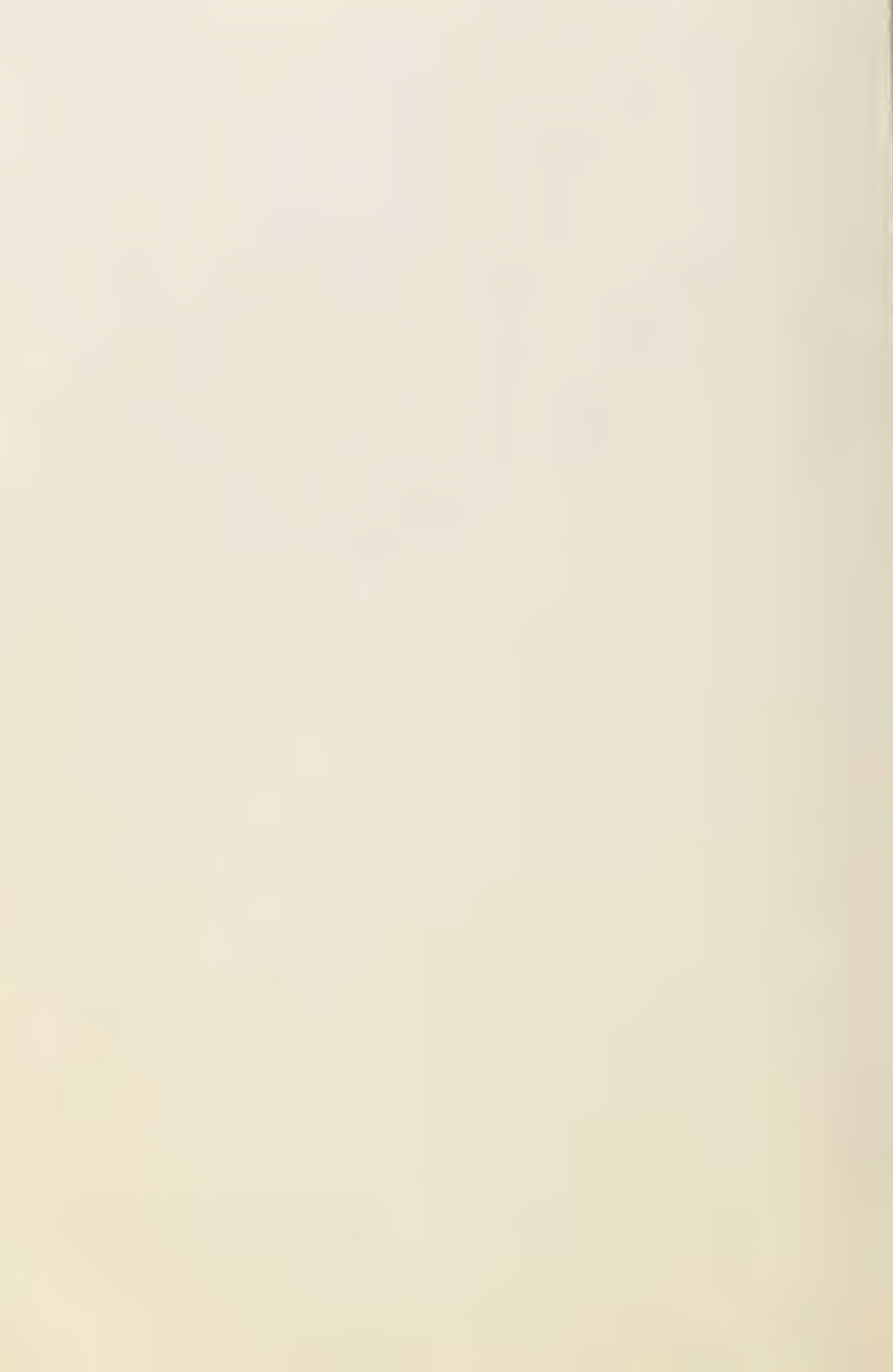
S. Weather Bureau, *Monthly mean sea level atmospheric pressure charts in the Northern*

Hemisphere, Extended Forecasting Sect. of U. S. Weather Bureau, Washington, D. C., published monthly, 1899-1958.

Wadsworth, G. P., Application of statistical methods of weather forecasting, *Compendium of Meteorol.*, Am. Meteorol. Soc., Boston, Mass., 849-855, 1951.

Wiener, N., *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, John Wiley & Sons, New York, 163 pp., 1949.

(Manuscript received August 17, 1959.)



Wave Propagation in a Medium with a Single Layer¹

L. KNOPOFF, F. GILBERT, AND W. L. PILANT

*Institute of Geophysics, University of California
Los Angeles, California*

Abstract—The generation of Love (SH) waves by an impulsive source located in a medium with a single layer has already been studied. The fine structure of the dispersive Love wave train has been shown to correspond to the superposition of a number of multiple reflections within the surface layer. The propagation of P-SV motions in a layer can be represented in a similar way as the sum of contributions from an infinite set of images of the source taken in the two boundaries. A seismic model has been set up to demonstrate the features of the wave propagation. The train of Rayleigh waves is observed to lengthen as the epicentral distance increases. The other important events are the refracted P and S pulse groups; each of these lengthens with increase in epicentral distance. The duration and the envelopes of the refracted P and S events are features not heretofore anticipated. A theoretical investigation based on a quantitative ray theory verifies the existence of these events. The quantitative ray theory is a method by which a number of the problems arising in seismology can be treated without the usual involved mathematical operations. An experimental technique for directly determining the group velocity in a model layered structure is presented. The results of the seismic model agree with the theoretical values.

Introduction—The solution to the problem of the excitation of Love waves in a medium with a single layer has already been given [Knopoff, 1958]. In the work referred to, the geometry considered is that of an elastic, homogeneous, isotropic, semi-infinite half-space with a homogeneous, elastic layer of finite thickness and different properties overlying it. In the upper layer, a line source is placed and allowed to vibrate impulsively parallel to its own length. If β_1 and β_2 are the shear velocities in the upper and the lower media, respectively, then Love waves are generated if $\beta_2 > \beta_1$.

Although the motion is exceedingly complex, such simplification arises if we compute only the first motions and convolute with a simple pulse. In this case, the results for the special case of the source located at $\frac{1}{2}d$ from the upper surface, density contrast $\rho_2/\rho_1 = 1$, and shear velocity contrast $\beta_2/\beta_1 = 2^{1/2}$ are presented in figure 1 for a number of stations in coordinates measured from the epicenter.

The properties of the motion are multifold, but the significant features of the motion are the following. Within the critical angle the re-

sponse assumes the form of a small number of discrete multiple reflections at steep angles. Once the first refraction has appeared, the remaining multiple reflections are so small in amplitude as to be insignificant. The duration of the Love wave train increases and the number of oscillations within the group increases as the epicentral distance increases. In addition to the properties of this wave train, derived on the basis of a steepest descent or other smoothing procedure valid at large distances, a fine structure to the wave train is obtained from this analysis. The major terms in the Love wave train are the events which arrive with the velocity β_1 and which correspond to multiple reflections within the upper layer; the refraction events arrive earlier than the corresponding reflection events but are not as strong as the reflection events for the short input pulse width considered.

In this paper it is proposed to study the P and SV motions of a medium with a single layer by methods similar to those considered above for the SH case. The two studies taken together should form a complete analysis of the motion in a medium with a single layer. Instead of exciting the medium at the center of the upper layer, we now excite the medium at

¹ Publication No. 152, Institute of Geophysics, University of California. This research was supported by the American Petroleum Institute.

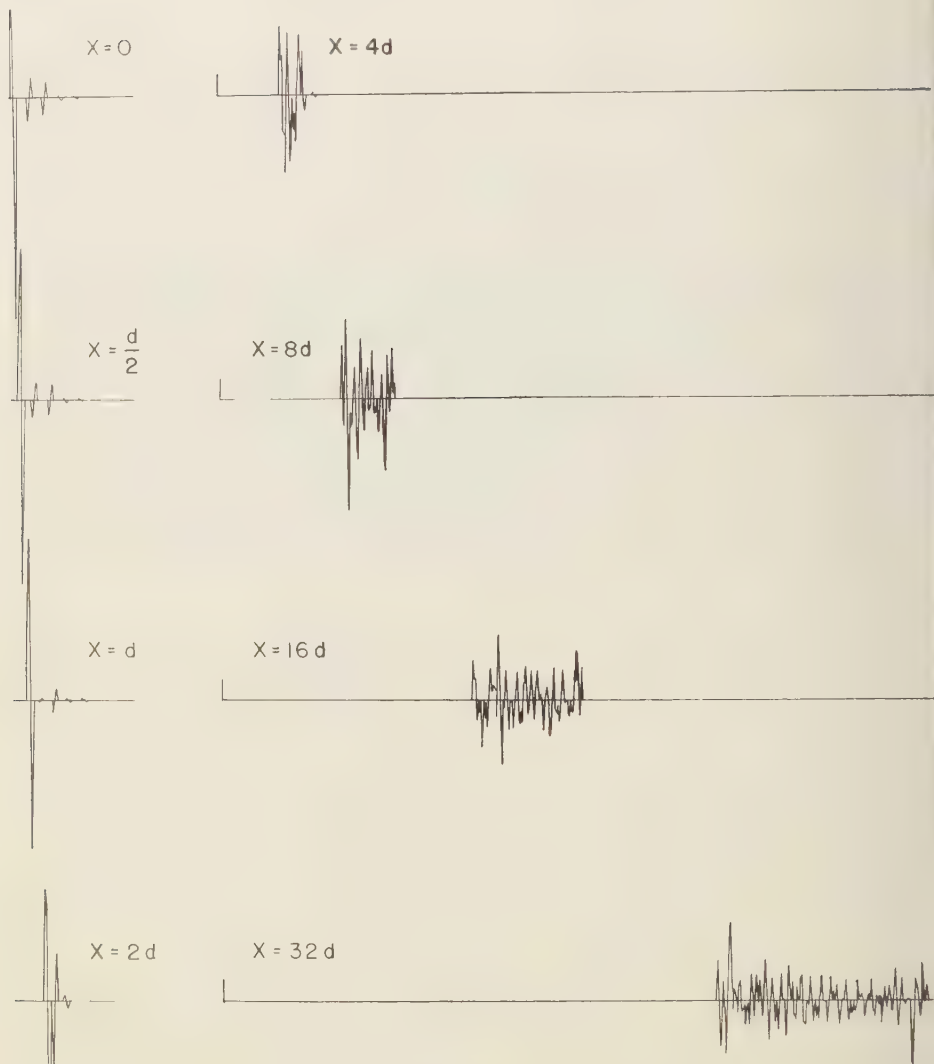


FIG. 1—Love waves from a pulsed line source located in the upper layer.

the upper, free surface. No significant differences in the results are expected from this change.

We propose to excite a uniform medium with a single layer by means of a single, isolated force directed normal to the upper surface and impulsive in time. Before proceeding to the mathematical analysis pertinent to our technique of solving this problem, we first review

the procedures used by other authors for a tack upon this problem and secondly report the results of some laboratory investigations upon a seismic model for this geometry.

Summary of previous work—Two significant pieces of work come to mind in reviewing the efforts of other authors to solve the problem of the excitation of a medium with a single layer by point, impulsive sources in the P-SV pola

tion. The first of these is the work of *Newlands* [1952] who has treated the problem by the Sommerfeld technique of deforming the contours and approximating the branch-line integrals by their values in the vicinity of the branch points. The treatment suffers from the usual difficulties associated with this method of approximation, and results are obtained which are valid only at large distances and for times near the arrivals.

The method was selected as an alternative to the ray procedure by *Newlands*, who said, 'It has long been recognized that the simple ray theory provides only a very incomplete picture of the disturbance at a point due to a sudden localized movement in an elastic medium.' We suggest that the solution to the problem can be obtained exactly by what is a substantial improvement in the 'ray theory,' although it is not a simple theory. All the features of the solution can be obtained by this improved technique.

The second treatment of this problem is the eigenvalue-eigenfunction analysis on which a large amount of work has been done for specific models. This research is summarized by *Ewing and others* [1957]. Unfortunately, it is not certain whether the solutions to the eigenvalue problem, as treated by these authors, may be superimposed to obtain the solution to the problem of the impulsive excitation of the medium with a layer, since the completeness of the eigensolutions has not been demonstrated. As pointed out by *Pekeris* [1948, p. 53] the equivalence of the mode solutions and the ray treatment has been demonstrated only for problems with simple, homogeneous boundary conditions. The mode solutions have been used to solve the problem of the impulsive excitation of a liquid layer overlying a liquid or a solid half-space. However these results are valid only for very large distances and do not explain refraction phenomena. Recently *Pekeris and others* [1959] have demonstrated, by an involved numerical procedure, that the low-frequency components in the ray theory, at sufficiently great distances, yield seismograms corresponding to those obtained by the normal mode theory; their solution was given for a liquid layer overlying a liquid half-space.

Both the preceding efforts are weighed down

with the burden of the involved algebra connected with the problem: first, sixth-order determinantal equations must be solved in order to find the reflection coefficients; second the manifold roots of the transcendental Rayleigh eigenequation for the surface wave velocities must be obtained, the equation itself having been derived from a sixth-order determinant.

In the later sections of this paper we shall consider the numerical aspects of the calculations as being technically difficult but conceptually simple. We shall concentrate our attention upon the procedures used and avoid the disguises offered by involved algebra.

Seismic model experiments, group velocity—A two-dimensional model of a half-space with a single layer has been constructed. The model consists of a 1/16-inch cold-rolled steel plate, to one edge of which a 1/16-inch \times 1/4-inch brass strip is silver soldered. The physical properties of the brass and steel are

Brass: $\alpha = 3.86 \text{ mm}/\mu\text{sec}$

$\beta = 2.21 \text{ mm}/\mu\text{sec}$

Steel: $\alpha = 5.34 \text{ mm}/\mu\text{sec}$

$\beta = 3.24 \text{ mm}/\mu\text{sec}$

The P wave velocities listed above are the plate velocities [*Ewing and others*, 1957, p. 283] and not the velocities for an infinite medium.

A source transducer of the usual type was placed on the model at the center of one of the long edges of the brass strip. A receiver was placed on the same outer edge at a fixed distance from the source (Fig. 2). The source transducer is excited by a sinusoidal pulse burst,

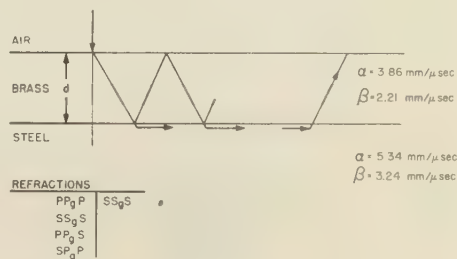


Fig. 2—Schematic diagram of refractions in a medium with a single layer.

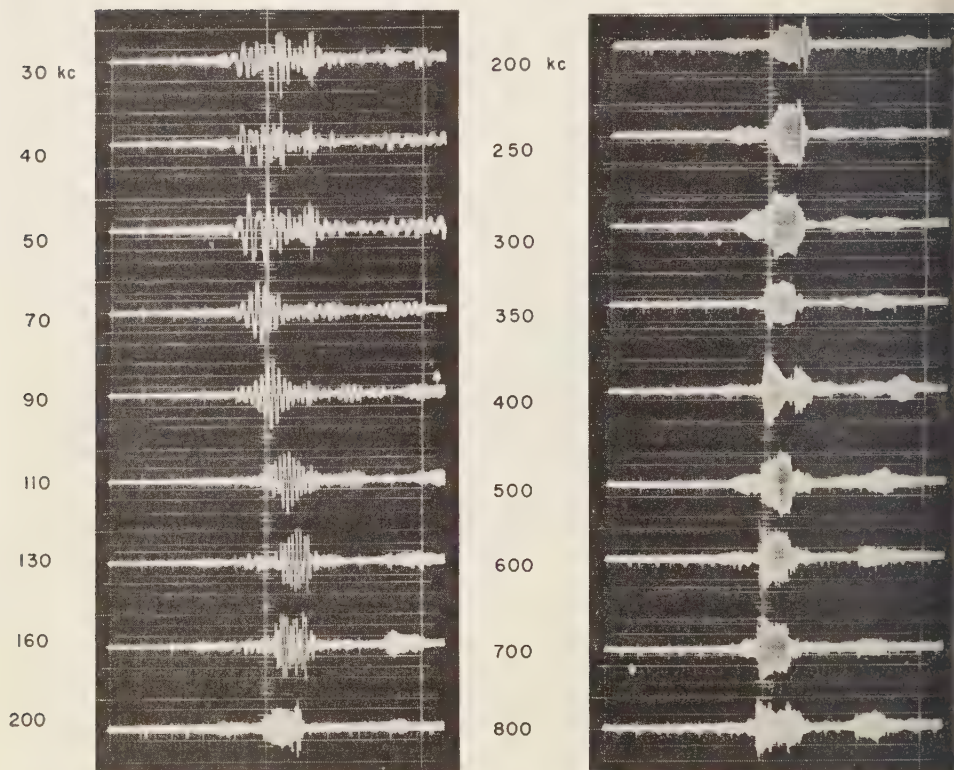


FIG. 3—Experimental group velocity dispersion curve as a function of carrier frequency.

a high-frequency carrier with a finite duration of several cycles. The burst duration is $60 \mu\text{sec}$ for all carrier frequencies. As the pulse group is transmitted through distance in the layered medium, the received signal no longer has a square envelope but rather has an envelope with a peak. By definition, the arrival time of the maximum in the modulation envelope is the arrival time of the group with the frequency of the carrier, and it thus determines the group velocity at this frequency.

The amount of energy that can be coupled into the model and the amount of detected signal depend upon the size of the transducer. When very small transducers are used almost no low-frequency energy can be detected. For this reason, to cover as wide a portion of the frequency band as possible, several transducers, both transmitters and receivers, of varying sizes were used. A composite run at several carrier

frequencies, taken with several transducers, is shown in Figure 3.

Even with great care in the selection of transducers, because of the small number of these, some records showed no marked peak in the modulation envelope. For all records a 'center of gravity' technique was used to determine the arrival of the group.

The sixth-order determinantal eigenequation has been solved for the Rayleigh phase velocity for the given velocities of brass and steel according to the standard technique given by Ewing, Jardetzky, and Press. This phase velocity curve is then differentiated in the usual way to obtain the group velocity curve. These curves and the experimental data are shown in Figure 4. The several types of marks indicating experimental points represent the use of different pairs of transducers. The data are seen to have included the group velocity mini-

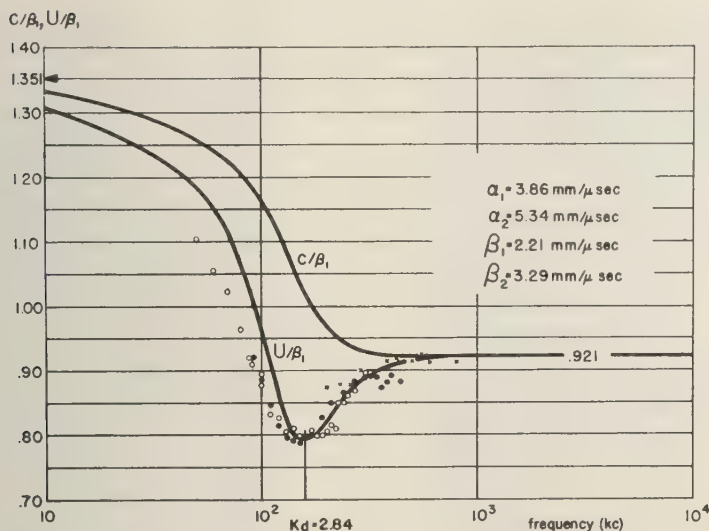


Fig. 4—Theoretical group and phase velocity curves for a brass layer overlying a steel half-space. Comparison is made with experimental values of group velocity obtained as in the method of Figure 3.

num associated with the Airy phase. This minimum is graphically observed in Figure 3 at a frequency of 160 kc/s.

Inaccuracies in the group velocity measurements are quite noticeable at the lower-frequency end of the data. This error arises because the determination of a modulation envelope peak is difficult in a frequency range where the peaks of individual sine waves are few and far apart. An error on the order of 10 per cent in the determination of the peak is not unreasonable.

Seismic model experiments, pulse excitation—The model of the preceding section was excited by a pulsed source-transducer placed at the free, brass surface of the bimetallic sheet. The excitation was of a conventional type, so that the stress applied to the surface was of very small spatial extent, the force was normal to edge, and the pulse duration was very short. The stress pulse was essentially an oscillatory pulse having a small number of extrema. The pulse had negligible tail and the pulse period was about 3.5 μ sec.

The receiving transducer was moved in 2-cm steps from 1 cm to 20 cm from the source and then by 5-cm steps to a distance of 50 cm along the brass edge. The detector responds to the nor-

mal component of the motion. The transducer system is now broadbanded in comparison with the transducer system used in the group velocity measurements.

The results of the measurements are given in Figure 5. The strongest signal is that of the Rayleigh wave, which is seen to be a train of oscillations; the duration of the train increases and the number of oscillations within the group increases as the epicentral distance increases, much as in the case of Love wave predictions for the same geometry. This indicates that the method of images will also apply to the present polarization.

Dispersion is to be seen in the Rayleigh wave train. In this case the trailing edge of the Rayleigh group has a longer period than the leading edge. This particular form of dispersion arises because most of the energy in the source is mainly concentrated in the high frequencies, frequencies higher than that for which the group velocity has a minimum.

In addition to the Rayleigh pulse, two other pulse groups are to be seen, both occurring earlier than the Rayleigh wave. In order to study these more closely the oscilloscope gain was increased, throwing the peaks of the Rayleigh waves off the screen but amplifying the

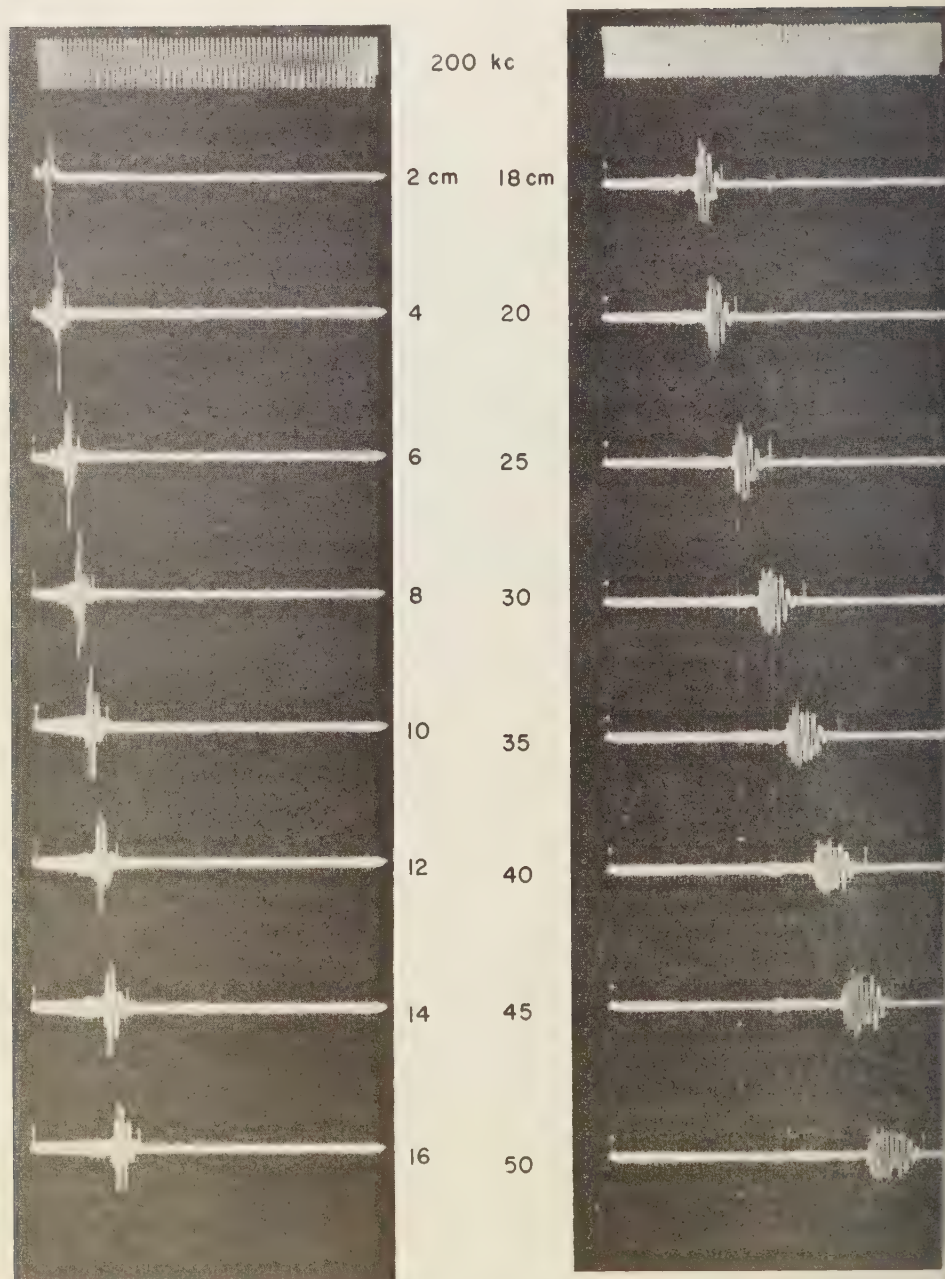


FIG. 5 Low-gain response of a pulsed medium with a single layer as a function of transducer separation.

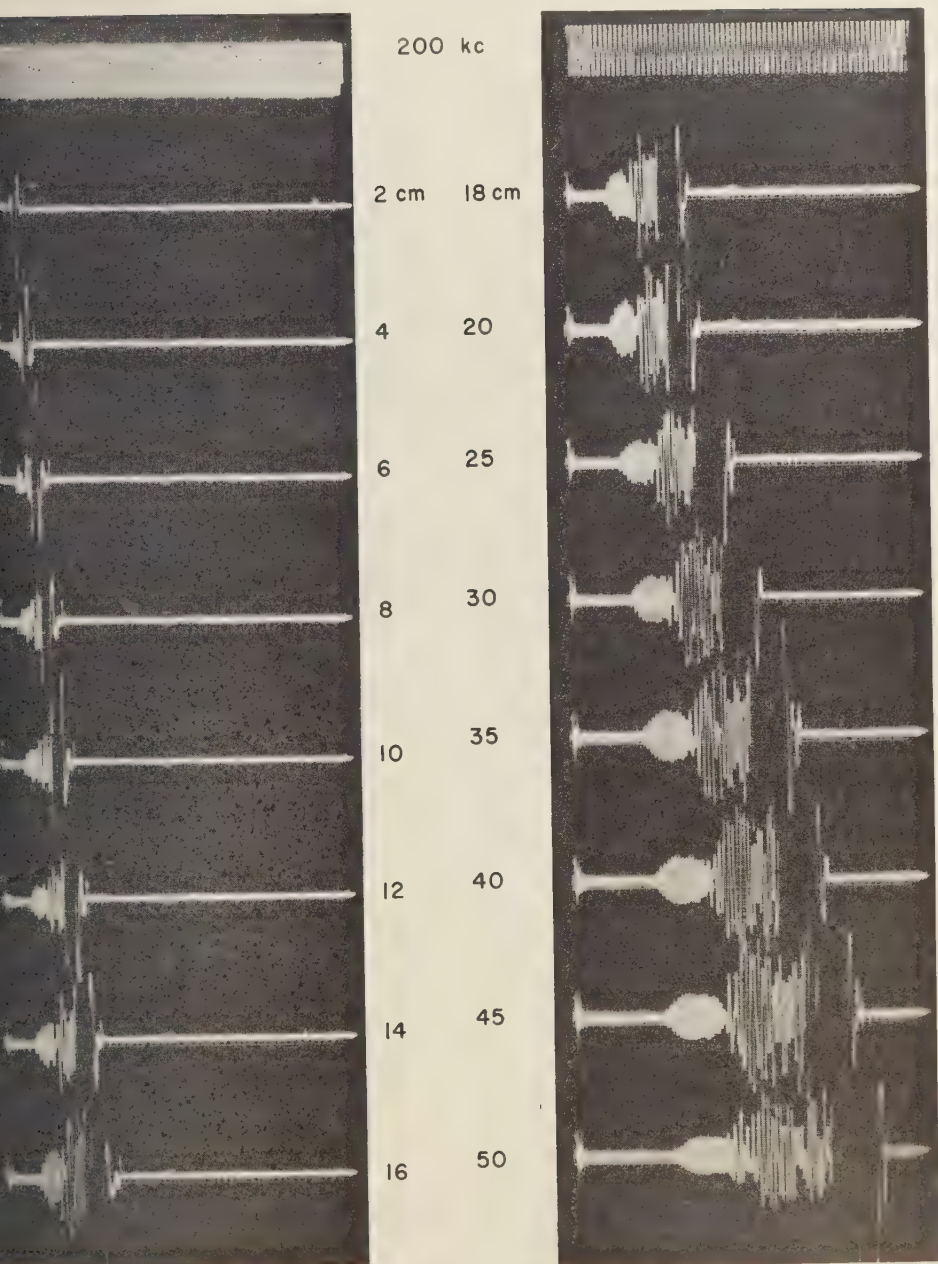


Fig. 6—High-gain response of a pulsed medium with a single layer as a function of transducer separation.

two precursor waves so that they became clear. An additional run at the same stations as before, but at this new gain, was taken, and these data are presented in Figure 6. An outstanding feature is the regular nature of the envelope to the first pulse burst.

The two pulse bursts are identified by travel-time seismic methods as P and S refractions through the steel plate. In view of the weakness of the refraction arrival in the Love wave prediction of Figure 1, the strength of the corresponding refraction arrivals in the present case must be explained. The reason is quite simple. We expand the source in an infinite series of images taken in the two surfaces of the upper layer. Each image can signal to the point of observation by a complex wave train consisting of a function which has a discontinuous value at the arrival of the reflections and which has a discontinuous first derivative at the arrival of the several refractions. In addition there is the discontinuity associated with the Rayleigh wave. Between these events the wave motion varies slowly and smoothly. According to the hypothesis of first motions, it is only the behavior at the arrivals that contributes to the response to an input pulse of short duration. Since the refraction arrivals integrate the responses of the reflection arrivals, it appears that they will be more significantly excited at low frequencies and should become less important at high frequencies of excitation. The critical time for comparison is the travel time through the upper layer. It was noted that in Figure 1 the excitation was assumed to be a pulse whose duration is short compared with the time $2d/\beta_1$. In the present case the time $2d/\alpha_1$ is about $3 \mu\text{sec}$; the period of the excitation is not small compared with this value and the refraction events should not be weak.

Love wave refractions—The theoretical problem of the prediction of the wave motion resulting from the surface excitation of the single layer is, in any practical case, at least an algebraically complicated one. We note that the first two principal pulse groups observed in the previous section are refractions through the higher-velocity region lying beneath. We concentrate our theoretical attack upon the refraction problem alone. Even here the problem is not algebraically simple.

In order to simplify the mechanics of the theoretical construction, we reconsider the theoretical results obtained for the refraction events in the problem of the Love wave motions in a medium with a single layer. In the problem a line SH source was placed at a distance $\frac{1}{2}d$ from the surface of a medium with a single layer. The layer has thickness d , shear velocity β_1 , and density ρ_1 . The lower medium has a shear velocity β_2 and density ρ_2 . The differential equation to be satisfied by the Laplace transform of the horizontal component of the motion is

$$\frac{\partial^2}{\partial x^2} \bar{U}_{1,2} + \frac{\partial^2}{\partial z^2} \bar{U}_{1,2} = \frac{s^2}{\beta_{1,2}^2} \bar{U}_{1,2} - \delta\left(x, z - \frac{d}{2}\right) F(s) \quad (1)$$

where $F(s)$ is the Laplace transform of the time dependence of the source. The motion due to the source alone, that is, if placed in an infinite medium, is

$$\bar{U}_1 = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{F(s)}{2\eta_1} \cdot \exp s\{px - \eta_1 |z - d/2|\} dp \quad (2)$$

where $\eta_{1,2} = (\beta_{1,2}^{-2} - p^2)^{1/2}$. The origin is taken at the interface between the two media. The transform of the motion at the surface can be shown to be represented by the integral

$$\bar{U} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{F(s)}{\eta_1} \exp s\{px - \eta_1(d/2)\} \cdot \left[\frac{1 + Qe^{-\eta_1 d}}{1 - Qe^{-\eta_1 d}} \right] dp \quad (3)$$

where

$$Q = \frac{1 - (\eta_2/\eta_1)(\rho_2\beta_2^2/\rho_1\beta_1^2)}{1 + (\eta_2/\eta_1)(\rho_2\beta_2^2/\rho_1\beta_1^2)} \quad (4)$$

We note that $|Q \exp(-\eta_1 d)| < 1$, and thus we may write

$$\bar{U} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{F(s)}{\eta_1} \exp s\{px - \eta_1(d/2)\} \cdot [1 + Qe^{-\eta_1 d}] \sum_{n=0}^{\infty} Q^n e^{-2n\eta_1 d} dp \quad (5)$$

Each term in the sum is the contribution to the total motion from the members of the finite set of images of the original source taken in the two planes, $z = 0$ and $z = d$.

At this point a mild digression in terms of plane wave theory is in order. We note that a plane SH wave incident at an angle θ from the side 1 upon a plane interface between two media of properties β_1 , ρ_1 and β_2 , ρ_2 may be presented by its Laplace transform

$$F(s) \exp \{s\beta_1^{-1}(x \sin \theta - z \cos \theta - \beta_1 t)\}.$$

The Laplace transform of the reflected wave is

$$AF(s) \exp (s/\beta_1)(x \sin \theta + z \cos \theta - \beta_1 t)$$

where the reflection coefficient A is, from the usual boundary conditions,

$$A = -\frac{1 - (\rho_1\beta_1/\rho_2\beta_2)(\cos \theta/\cos \phi)}{1 + (\rho_1\beta_1/\rho_2\beta_2)(\cos \theta/\cos \phi)} \quad (6)$$

where ϕ is the angle of emergence of the wave into medium 2,

$$\sin \theta/\sin \phi = \beta_1/\beta_2$$

It is thus evident that the plane wave representation of the source function

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{F(s)}{\eta_1} \exp s\{px - \eta_1(d/2)\} dp$$

the plane wave representation of the reflection coefficients

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} Qe^{-\eta_1 d} dp$$

or the case of reflections from the interface between the media, and the plane wave representation of the reflection coefficients for the reflections from the free surface

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{-\eta_1 d} dp$$

may all be combined into a complete representation of the motion. We write (5), not by solving the boundary value problem, but by combining the source function and the appropriate reflection coefficients by inspection of the geometry.

The typical integral

$$I_{i,m}^0 = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} Q^i \frac{F(s)}{\eta_1} \exp s(px - m\eta_1 d) dp \quad (7)$$

has already been evaluated, but the technique is of importance and thus we shall briefly review the procedure. In this equation $m = n + \frac{1}{2}$ where n is a positive integer. We make the change of variable

$$-t = px - m\eta_1 d \quad (8)$$

where s and t are taken as real variables. There are branch points located at $\pm \beta_1^{-1}$, $\pm \beta_2^{-1}$ in the p -plane, and we draw cuts in the p -plane from each branch point to infinity.

The original contour has been deformed into a hyperbola with vertex at $-\sin \theta/\beta_1$. Now for the refraction events $\sin \theta > \beta_1/\beta_2$, the vertex of the hyperbola lies on the cut between the two branch points $-1/\beta_1$ and $-1/\beta_2$. We thus deform the contour of integration about the end of the cut at $-1/\beta_2$ (Fig. 7).

The value of t at the branch point $-1/\beta_2$ is the time

$$t_0 = (r/\beta_2) \{\sin \theta + [(\beta_2/\beta_1)^2 - 1]^{1/2} \cos \theta\}$$

the time of arrival of the refraction event.

Solving (8) for $p(t)$, we obtain for the straight-line portions of the path.

$$p = -(t/r) \sin \theta \quad (9)$$

$$+ (t/r)[(r/\beta_1 t)^2 - 1]^{1/2} \cos \theta$$

where $\sin \theta = x/r$ and $r^2 = x^2 + (md)^2$.

The integral of (7) over the straight-line portions of the path in Figure 7 is

$$I_{i,m}^0 = Im \frac{1}{\pi} \int_{-1/\beta_2}^{-(1/\beta_1) \sin \theta} Q^i(p) \frac{F(s)}{\eta_1} \exp s(px - m\eta_1 d) dp \quad (10)$$

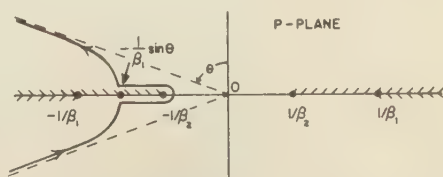


Fig. 7—Deformed path of integration in the p -plane for reflected events in the Love wave problem.

where $Q(p)$ is now a complex number due to the behavior of η_2 for $\text{Re}\{p\} < -1/\beta_2$. The values of $Q(p)$ on the two sides of the cut are complex conjugates. Changing the variable of integration in (7) to t by means of (8), we obtain

$$I_{i,m}^0 = Im \frac{1}{\pi} \int_{t_0}^{r/\beta_1} e^{-st} \cdot \frac{Q^i(p)}{(\beta_1^{-2} - p^2)^{1/2}} \frac{dp}{dt} dt \quad (11)$$

where we take the upper sign in (9) for the values of $p(t)$ in the interval of integration.

The integral in (11) is the Laplace transform of

$$\frac{1}{\pi} Im\{Q^i(p)\} \frac{1}{(\beta_1^{-2} - p^2)^{1/2}} \frac{dp}{dt} \cdot [H(t - t_0) - H(t - r/\beta_1)] \quad (12)$$

Now from (9),

$$\frac{1}{(\beta_1^{-2} - p^2)^{1/2}} \frac{dp}{dt} = -\frac{1}{[(r/\beta_1)^2 - t^2]^{1/2}} \quad (13)$$

Thus the motion at the surface for the refraction events depends only upon the properties of the reflection coefficients Q and (12) becomes

$$\frac{1}{\pi} Im\{Q^i(p)\} \frac{1}{[(r/\beta_1)^2 - t^2]^{1/2}} \cdot [H(t - t_0) - H(t - r/\beta_1)] \quad (14)$$

Now $Im\{Q(p)\}$ vanishes at $p = -1/\beta_2$, the tip of the lancet, corresponding to the arrival of the refraction, since $\eta_2 = 0$ at this point. Letting $p = -1/\beta_2 + dp/dt|_{t_0}(t - t_0)$ for points near the refraction event, we have

$$\eta_2 \approx i\{(2/\beta_2)(t - t_0)\eta_1 \cdot [(r/\beta_1)^2 - t_0^2]^{-1/2}\}^{1/2} \quad (15)$$

where we have applied (13). Thus

$$Im\{Q^i(p)\} = -2j \frac{\rho_2 \beta_2^2}{\rho_1 \beta_1^2} \left(\frac{2}{\beta_2}\right)^{1/2} \cdot \frac{(t - t_0)^{1/2}}{[(r/\beta_1)^2 - t_0^2]^{1/4} [\beta_1^{-2} - \beta_2^{-2}]^{1/4}} \quad (16)$$

and thus

$$I_{i,m}^0 = -\frac{2}{\pi} j \frac{\rho_2 \beta_2^2}{\rho_1 \beta_1^2} \left(\frac{2}{\beta_2}\right)^{1/2} \cdot \frac{(t - t_0)^{1/2}}{[(r/\beta_1)^2 - t_0^2]^{3/4} [\beta_1^{-2} - \beta_2^{-2}]^{1/4}} \quad (17)$$

for times near the refraction event.

P-SV Refractions—We now turn to the mathematical analysis of the wave motion which results from the surface excitation of a medium with a single layer. The geometry is as before the properties of the layer are $\alpha_1, \beta_1, \rho_1$ and of the half-space $\alpha_2, \beta_2, \rho_2$. The free surface of the layer is excited by a line source exerting the component τ_{xx} and having a time dependence $\delta(t)$ at the origin.

By reference to the preceding section we can immediately write the contour integral for any given ray path of a refracted wave. Consider the ray which has undergone the following reflections of given type at the upper surface.

- ν reflections of type P-S
- μ reflections of type S-P
- ρ reflections of type S-S
- κ reflections of type P-P

In addition, let the wave have been emitted as a P wave from the source and received as a P wave at the receiver. Further let there have been taken place

- n reflections of type P-S
- m reflections of type S-P
- r reflections of type S-S
- q reflections of type P-P

at the lower surface. Then the vertical component of motion at the free surface has a Laplace transform given by the contour integral over the usual contour

$$\begin{aligned} \bar{U}_z = & \frac{1}{2\pi i} \int_{\Gamma} \exp\{spx\} \\ & - (\nu + \mu + 2\kappa + 2)s\eta_{P_1} d \\ & - (\nu + \mu + 2\rho)s\eta_{s_1} d\} dp \\ & \cdot S_P(p)(R_{PP}^+)^{\nu}(R_{PS}^+)^{\mu}(R_{SP}^+)^{\kappa}(R_{SS}^+)^{\rho} \\ & \cdot (R_{PP}^-)^{\nu+\kappa+1-m}(R_{PS}^-)^{m+\mu+\rho} \\ & \cdot (R_{SP}^-)^m(R_{SS}^-)^{\nu+\rho-m} D_P(p) \end{aligned} \quad (18)$$

where R^+ are the plane wave reflection coefficients for the waves of appropriate type at the upper surface, R^- are the corresponding plane wave reflection coefficients at the lower surface, $\nu(p)$ is a source distribution function for the P wave part of the source and $D_P(p)$ is a receiver directivity function giving the vertical component of the motion resulting from the incidence of a P wave upon the free surface. We state the following relationships.

$$\nu - \mu = m - n$$

$$\kappa - \rho + 1 = q - r$$

$$+ \nu + \kappa + \rho + 1 = m + n + r + q \quad (19)$$

The source functions are, from an elementary theory [Lamb, 1904],

$$S_P(p) = \frac{(2p^2 - \beta_1^{-2})}{\mu s R(p)} \quad (20)$$

$$S_S(p) = -\frac{2p\eta_{p1}}{\mu s R(p)}$$

where the Rayleigh denominator $R(p)$ is

$$R(p) = (2p^2 - \beta_1^{-2})^2 + 4p^2\eta_{p1}\eta_{s1} \quad (21)$$

and we define $\eta_p^{-1} = (\alpha_1^{-2} - p^2)^{1/2}$, etc. Again, from elementary theory,

$$R_{PS}^+(p) = \frac{-(2p^2 - \beta_1^{-2})^2 + 4p^2\eta_{p1}\eta_{s1}}{R(p)} \quad (22)$$

$$R_{PS}^+(p) = \frac{4p\eta_{p1}(2p^2 - \beta_1^{-2})^2}{R(p)} \quad (23)$$

$$R_{SS}^+(p) = R_{PP}^+(p) \quad (24)$$

$$\begin{aligned} R_{PS}^+(p) &= \frac{-\eta_{s1}}{\eta_{p1}} R_{PP}^+(p) \\ &= \frac{-4p\eta_{s1}(2p^2 - \beta_1^{-2})^2}{R(p)} \end{aligned} \quad (25)$$

The directivity functions [Knopoff and others, 1957] are

$$D_P(p) = \frac{-\beta_1^{-2}\eta_{p1}(2p^2 - \beta_1^{-2})^2}{R(p)} \quad (26)$$

$$D_S(p) = \frac{-4p\eta_{p1}\eta_{s1}\beta_1^{-2}}{R(p)}$$

The appropriate changes in (18) and in the compatibility relations (19) may be made for combinations of emitted P or S waves, and received P or S waves.

The reflection coefficients R^- are obtained by standard procedures. These involve the solution of four simultaneous equations. The reflection coefficients will not be written explicitly here, but it will be assumed that these are obtainable. As we shall see these reflection coefficients need be evaluated only at the critical angle.

Let it be assumed that we are interested in refractions that travel with the P wave velocity in the lower medium. For instance the event PP_gP corresponds to the conditions $\nu \doteq \mu = \rho = \kappa = 0$, $n = m = r = 0$, and $q = 1$ where we use standard seismic notation for the refraction P_g . The path in the p -plane is given by

$$\begin{aligned} t(p) &= px - (\nu + \mu + 2\kappa + 2)\eta_{p1}d \\ &\quad - (\nu + \mu + 2\rho)\eta_{s1}d \end{aligned} \quad (27)$$

This path is not a simple hyperbola in the cases where the coefficients of $\eta_{p1}d$ and $\eta_{s1}d$ are both nonvanishing. The P_g refractions arise from diverting the contour $p(t)$ about the branch point at $p = -1/\alpha_2$ and evaluating the contour integral over the two straight-line portions of the path in the vicinity of this branch point. The value of (27) at the branch point $t^0 = t(-1/\alpha_2)$ is the arrival time of the refraction.

The refraction event thus has an amplitude

$$U_z = \frac{1}{\pi} \left(\frac{dp}{dt} \right)^0 D_P^0 S_P^0 \quad (28)$$

$$\begin{aligned} &\cdot (R_{PP}^{+0})^{\kappa} (R_{PS}^{+0})^{\nu} (R_{SP}^{+0})^{\mu} (R_{SS}^{+0})^{\rho} \\ &\cdot I m \{ (R_{PP}^{-1})^{\nu+\kappa+1-m} (R_{PS}^{-1})^{m+\mu-\nu} \\ &\cdot (R_{SP}^{-1})^m (R_{SS}^{-1})^{\nu+\rho-m} \} \end{aligned}$$

where $(dp/dt)^0$ is the value of this quantity obtained from (27). All superscripts 0 imply evaluation of these quantities at $p = -1/\alpha_2$. The superscripts 1 indicate that at $p = -1/\alpha_2$ the quantity $\eta_p = 0$ and an expansion of the type made previously (15) must be constructed.

$$\eta_{p2} = i \left\{ \frac{2}{\alpha_2} (t - t_0) \left| \frac{dp}{dt} \right|^0 \right\}^{1/2} \quad (29)$$

TABLE 1—Relative amplitudes of the first 84 terms in the series representing the P_g refraction event

n^*	$P_n S_{2-n}$	$P_n S_{4-n}$	$P_n S_{6-n}$
0	0.0963	0.03431	-0.01361
1	1.0000	-0.2268	0.03231
2	-0.9340	1.8276	-0.2882
		-2.1411	2.8730
		0.2732	-5.0550
			0.9165
			-0.1206

* n denotes the number of traversals of the layer as a P wave. For example, from the $n = 2$ entry, the $P_2 S_0$ event (a PP event) has an amplitude relative to the $P_1 S_1$ event of -0.9340, whereas the $P_2 S_2$ event (2P, 2S traversals including terms PPSS, PSSP, SPPS, SPSP, SSPP) has a relative amplitude of 1.8276. Note that $P_1 S_1$ includes the events $PS(P_g)$ and $SP(P_g)$.

where $(dp/dt)^0$ is again obtained from (27) and evaluated at $p = -1/\alpha_2$.

This is a very laborious procedure but the evaluation has been made for the first 84 P_g refraction events by this method. The 84 events are the events that travel in the upper layer as the simple P_g refractions PP, PS, SP, SS (corresponding to cases $\mu = \nu = \rho = \kappa = 0$ and one of $m, n, r, q = 1$, the others zero), the sixteen combinations of four traversals in the upper medium, for example, PPPP, PSPP, PPSP, SSPP, and the 64 combinations involving six traversals in the upper medium, for example, PPPPPP, PSSPPS, SSSPPS. For the velocities of brass and steel given above we obtain the amplitudes of these 84 terms and present them in Table 1, where the numerical values given are the coefficients of $(t-t^0)^{1/2}$ normalized to unity for the event $SP_g P + PP_g S$. The evaluation has been made at the point in the upper surface $x = 80d$.

In addition we have calculated the coefficients of $(t-t^0)^{1/2}$ for the events

$$\begin{aligned}
 P(PP)^N P(P_g) \quad \nu = \mu = \rho = 0 \\
 \kappa = N \\
 n = m = r = 0 \\
 q = N + 1 \quad (a)
 \end{aligned}$$

TABLE 2—Relative amplitudes of S, P, and symmetrically mixed S and P terms in the series representing the P_g refraction event

N	SS	PP	SP + PS
0	0.0963	-0.9340	1.0000
1	0.0344	0.2701	1.1606
2	0.00919	-0.0585	1.0096
3	0.00219	0.0113	0.7798
4	0.00048	-0.00204	0.5646
5			0.3923
6			0.2660
7			0.1749

The events are symmetric, that is, for the entry PP, N refers to the event $P(PP)^N P$ while SP, N refers to $S(PS)^N P$. N is the number of reflections from the top of the layer. For example, from the $N = 1$ entry, the relative amplitude of the event SSSSS is 0.00919, whereas the relative amplitude of the event SPSPSP + PSPSPS is 1.0096.

$$\begin{aligned}
 P(SP)^N S(P_g) \quad \nu = \kappa = \rho = 0 \\
 \mu = N \\
 n = q = r = 0 \\
 m = N + 1 \quad (b) \\
 S(PS)^N P(P_g) \quad \mu = \kappa = \rho = 0 \\
 \nu = N \\
 m = q = r = 0 \\
 n = N + 1 \quad (c) \\
 S(SS)^N S(P_g) \quad \mu = \kappa = \nu = 0 \\
 \rho = N \\
 m = q = n = 0 \\
 r = N + 1 \quad (d)
 \end{aligned}$$

for values of N from 0 to 7, where the symbol (P_g) indicates that these reflections in the upper medium are coupled to a P_g refraction in the lower medium. The numerical values for the amplitudes are presented in Table 2 for $x = 80d$ and plotted in Figure 8. A continuous curve has been drawn through the numbers as a function of N but it is understood that only integral values of N are significant. The numbers have been normalized to the same value as before. The events (b) and (c) above arrive simultaneously for the same N and thus the sum of their amplitudes has been plotted. The thr

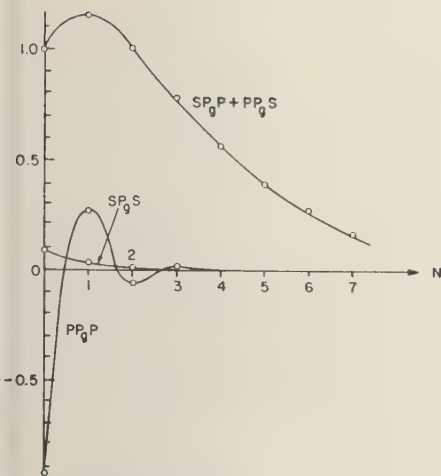


Fig. 8—Relative amplitudes of the events $(PS)^{N+1}(P_g)$, $(PP)^{N+1}(P_g)$, $(SS)^{N+1}(P_g)$.

of events (a), (b)–(c), and (d) of course all arrive at different times even for the same N . The feature of interest in these numbers and those of Table 1 is the fact that most of the energy of the refraction arrivals is concentrated among the events that have a large number of PS reflections at the upper surface. The series of pulses may now be convoluted with the input pulse. We first consider the data of Table 2. We note the nature of the pulse over very short range; this is assumed to describe the convolution of the input stress pulse in time with the function $(t - \tau)^{1/2}$. This input function has been drawn in Figure 9b and for

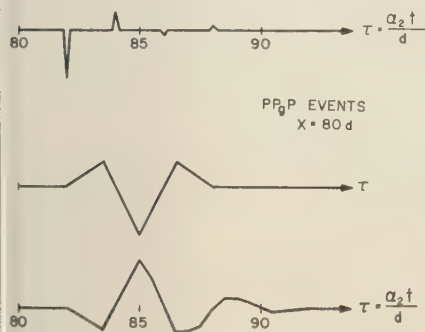


Fig. 9—Convolution 9c of the input pulse 9b with the short pulse response 9a.

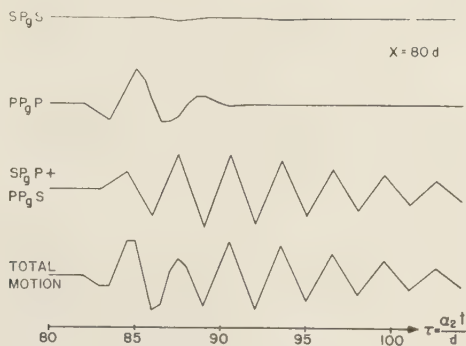


Fig. 10—Convolution of the input pulse with the events given in Figure 8.

convenience is represented by a series of straight lines. In order to convolute this function with a suite of simple pulses of the sequence (a) $P(PP)^N P(P_g)$ whose amplitudes and times are given in Figure 9a we merely repeatedly add together the input pulse with the appropriately scaled amplitudes occurring at the appropriate times. The result of this convolution is shown in Figure 9c. This technique gives the correct convolution result, since the single pulse response is $(t - t_0)^{1/2}$, the function assumed for the input stress even at the shortest range.

In a similar way the convolution with the other entries in Table 2 can be obtained, together with their sum for the given input function. These results are given in Figure 10. It is seen, again, that the major features of the response are connected with the waves $(SP)^N$ and $(PS)^N$, whereas the others are not as important except near the onset of the wave train.

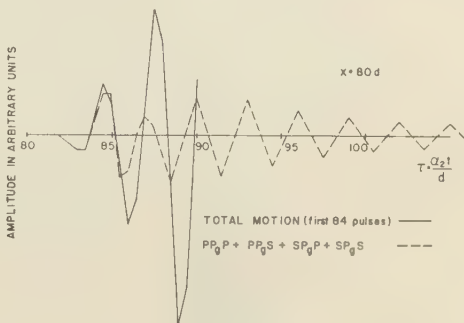


Fig. 11—Convolution of the input pulse with the first 84 refraction events.

The result thus obtained is a fair picture of the experimental result at $x = 80d$ (that is, $x = 50$ cm).

In order to determine the effect of omitting the other reflection combinations, we have convoluted the input pulse with *all* the first 84 events of amplitudes given in Table 1. This result is given in Figure 11 and is carried out only to the time that the first of the eight-fold traversals of the upper medium $P_s(P_g)$ arrives. The result of Figure 10 is included for comparison. The errors are significant. The result of Figure 11 is in better accord with the experimental observation.

REFERENCES

EWING, M., W. S. JARDETSKY, AND F. PRESS, *Elastic Waves in Layered Media*, McGraw-Hill, New York, 1957.

- KNOPOFF, L., Love waves from a line SH source, *J. Geophys. Research*, **63**, 619-630, 1958.
- KNOPOFF, L., R. W. FREDRICKS, A. F. GANGI, AND L. D. PORTER, Surface amplitudes of reflect body waves, *Geophysics*, **22**, 842-847, 1957.
- LAMB, H., On the propagation of tremors over the surface of an elastic solid, *Phil. Trans. Roy. Soc. London, Ser. A* **203**, 1-42, 1904.
- NEWLANDS, M., The disturbance due to a line source in a semi-infinite elastic medium with a single surface layer, *Phil. Trans. Roy. Soc. London, Ser. A*, **245**, 213-308, 1952.
- PEKERIS, C. L., Theory of propagation of explosive sound in shallow water, *Geol. Soc. Am., Memoir*, **27**, 53, 1948.
- PEKERIS, C. L., I. M. LONGMAN, AND H. LIFSCHITZ, Application of ray theory to the problem of long-range propagation of explosive sound in a layered liquid, *Bull. Seismol. Soc. Am.*, **49**, 242-250, 1959.

(Manuscript received July 30, 1959.)

Nonlinear Oscillations of a Particle in a Long Wave in a Rotating Fluid*

S.-K. KAO

*Department of Meteorology, University of California
Los Angeles, California*

Abstract. The effects of a long wave of finite amplitude on the oscillation of a particle in a rotating fluid are examined by solving the Lagrangian equations of motion. It is shown that the motion of the particle is nonoscillatory or oscillatory, depending respectively on whether the forcing frequency kUr is equal to the multiples of the natural frequency or not. In the former case, the zeroth-order solution consists of the natural and forcing frequencies, whereas the first-order solution gives three new frequencies. In the latter case, the zeroth-order solution consists of the natural frequency with linearly growing amplitude, whereas the first-order solution gives a high frequency oscillation of twice the natural frequency with linearly growing amplitude. A comparison of the frequencies predicted by the zeroth-order and first-order solutions and those calculated from the trajectories of 300-mb constant-pressure balloons is made.

Introduction. Analysis of oscillations of a particle in a pressure field of a rotating fluid has the direct implication that it predicts particle trajectories and leads to the understanding of the mechanism of diffusion in a rotating fluid. Mathematically, this involves integrating a set of Lagrangian equations of motion in which the pressure force is generally a function of space and time. In general, these differential equations are highly nonlinear. Only a few exact solutions have been found for some simple types of pressure fields [Forsythe, 1949; Kao and Wurtele, 1959].

Attempts have been made to investigate oscillations and trajectories of a particle in some model pressure systems which characterize the pressure distribution in a rotating fluid. The nonlinear oscillations and trajectories of a particle in a pressure field with a geostrophic profile have been analyzed by Kao and Wurtele [1959]. A method of linearizing the system of Lagrangian equations and estimating the zeroth-order oscillation of a particle in some types of pressure systems has been suggested by Kao and Neiburger [1959]. However, this method is valid only for pressure disturbances of infinitesimal amplitude. The objectives of this paper are to analyze the nonlinear characteristics and to seek an approximate solution for the

oscillations and trajectories of a particle in a pressure wave of finite amplitude in a rotating fluid.

Lagrangian equations of motion and method of approximation. We consider the oscillations of a particle in a horizontal plane in a rotating, inviscid fluid. Let x, y be the right-hand rectangular coordinates in the plane. As a particle moves in the plane its position is a function of time t : $x = x(t)$, $y = y(t)$. The time derivatives of these functions are components u, v of the particle velocity at the point (x, y) at time t : $u = u(t) = dx/dt$, $v = v(t) = dy/dt$. The Lagrangian equations of motion for a particle in a wave of the form $Uy + Ak^{-1} \cos k(x - ct)$ may be written

$$d^2x/dt^2 - f(dy/dt) = -fA \sin k(x - ct) \quad (1)$$

$$d^2y/dt^2 + f(dx/dt) = fU \quad (2)$$

where f denotes twice the angular velocity of the rotating system, k the wave number, c the phase velocity, and A and U are constants in the dimension of velocity. The above system of differential equations is highly nonlinear. Because this system is a nonautonomous one, all concepts such as singular points and limit cycles become useless. Not only is explicitly solving this system of equations beyond the frontiers of mathematical knowledge [Poincaré, 1892; Van der Pol, 1926; Kryloff and Bogoluboff, 1943; Andronow and Chaikin, 1949; Minosky, 1946; Lejscheltz, 1946; Stoker, 1950], but even the advantages of

*Publication No. 48, Department of Meteorology, UCLA. This work was performed under the auspices of Naval Research Contract Nonr-233(21).

a topological representation in the phase plane disappear.

We introduce

$$\begin{aligned}\xi &= x(t) - Ut - x(0) \\ \eta &= y(t) - y(0) \quad \tau = ft\end{aligned}\quad (3)$$

For brevity, let the initial conditions $x(0)$, $y(0)$, $u(0)$, and $v(0)$ be respectively denoted by α , β , u_0 , and v_0 . Furthermore, let

$$U_\tau = U - C \quad \lambda = U_\tau/f \quad a = A/f$$

Equations (1) and (2) may be written

$$d^2\xi/d\tau^2 - d\eta/d\tau = -a \sin k(\lambda\tau + \xi + \alpha) \quad (4)$$

$$d^2\eta/d\tau^2 + d\xi/d\tau = 0 \quad (5)$$

The converted initial conditions at $\tau = 0$ are $\xi = 0$, $\eta = 0$, $d\xi/d\tau = (u_0 - U)/f$, and $d\eta/d\tau = v_0/f$.

Expanding the right-hand side of (4) into Taylor's series, we obtain

$$\begin{aligned}\frac{d^2\xi}{d\tau^2} - \frac{d\eta}{d\tau} &= -a \sin k(\lambda\tau + \alpha) \\ &\cdot \left\{ 1 - \frac{1}{2!} (k\xi)^2 + \frac{1}{4!} (k\xi)^4 - + \dots \right\} \\ &- a \cos k(\lambda\tau + \alpha) \\ &\cdot \left\{ k\xi - \frac{1}{3!} (k\xi)^3 + \frac{1}{5!} (k\xi)^5 - + \dots \right\} \quad (6)\end{aligned}$$

The classical perturbation method of obtaining an approximate solution of a nonlinear differential equation is to expand the dependent variable in a power series of a small parameter and then to calculate the series term by term. However, nonlinear differential equations generally contain singularities at some locations in zeroth- and higher-order solutions. The power series expansion on a small parameter breaks down near singularities, and the classical perturbation fails to give a usable solution near the singular point. To eliminate such difficulties we shall follow Lighthill [1949] and expand both dependent and independent variables in power series of a small parameter. For sufficiently long waves, we may consider k a small parameter and expand ξ , η , and τ as follows:

$$\xi(\gamma) = \xi_0(\gamma) + k\xi_1(\gamma) + k^2\xi_2(\gamma) + \dots \quad (7)$$

$$\eta(\gamma) = \eta_0(\gamma) + k\eta_1(\gamma) + k^2\eta_2(\gamma) + \dots \quad (8)$$

$$\tau(\gamma) = \gamma + k\tau_1(\gamma) + k^2\tau_2(\gamma) + \dots \quad (9)$$

where γ takes the place of the original independent variable τ . Evidently $\xi_0(\gamma)$ and $\eta_0(\gamma)$ are simply the zeroth-order solution of the classical method, with γ replacing τ .

Substituting (7), (8), and (9) in (6) and (5) we have respectively,

$$\begin{aligned}& \frac{\xi_0'' + k\xi_1'' + k^2\xi_2'' + \dots}{(1 + k\tau_1' + k^2\tau_2' + \dots)^2} - \frac{(\xi_0' + k\xi_1' + k^2\xi_2' + \dots)(k\tau_1'' + k^2\tau_2'' + \dots)}{(1 + k\tau_1' + k^2\tau_2' + \dots)^3} \\ & - \frac{\eta_0' + k\eta_1' + k^2\eta_2' + \dots}{1 + k\tau_1' + k^2\tau_2' + \dots} \\ & = -a \sin k(\lambda\gamma + \alpha) \left\{ \left[1 - \frac{1}{2!} k^2\lambda^2(k\tau_1 + k^2\tau_2 + \dots)^2 \right. \right. \\ & \quad \left. \left. + \frac{1}{4!} k^4\lambda^4(k\tau_1 + k^2\tau_2 + \dots)^4 - + \dots \right] \right. \\ & \quad \times \left[1 - \frac{1}{2!} k^2(\xi_0 + k\xi_1 + k^2\xi_2 + \dots)^2 + \frac{1}{4!} k^4(\xi_0 + k\xi_1 + k^2\xi_2 + \dots)^4 - + \dots \right] \\ & \quad - \left[k\lambda(k\tau_1 + k^2\tau_2 + \dots) - \frac{1}{3!} k^3\lambda^3(k\tau_1 + k^2\tau_2 + \dots)^3 + - \dots - \dots \right] \\ & \quad \times \left[k(\xi_0 + k\xi_1 + k^2\xi_2 + \dots) - \frac{1}{3!} k^3(\xi_0 + k\xi_1 + k^2\xi_2 + \dots)^3 + - \dots \right] \left. \right\}\end{aligned}$$

$$\begin{aligned}
 & -a \cos k(\lambda\gamma + \alpha) \left\{ \left[1 - \frac{1}{2!} k^2 \lambda^2 (k\tau_1 + k^2\tau_2 + \dots)^2 \right. \right. \\
 & \left. \left. + \frac{1}{4!} k^4 \lambda^4 (k\tau_1 + k^2\tau_2 + \dots)^4 - + \dots \right] \right. \\
 & \times \left[k(\xi_0 + k\xi_1 + k^2\xi_2 + \dots) - \frac{1}{3!} k^3 (\xi_0 + k\xi_1 + k^2\xi_2 + \dots)^3 + - \dots \right] \\
 & + \left[k\lambda(k\tau_1 + k^2\tau_2 + \dots) - \frac{1}{3!} k^3 \lambda^3 (k\tau_1 + k^2\tau_2 + \dots)^3 + - \dots \right] \\
 & \times \left[1 - \frac{1}{2!} k^2 (\xi_0 + k\xi_1 + k^2\xi_2 + \dots)^2 + \frac{1}{4!} k^4 (\xi_0 + k\xi_1 + k^2\xi_2 + \dots)^4 - + \dots \right] \left. \right\} \\
 & \quad \quad \quad (10)
 \end{aligned}$$

and

$$\begin{aligned}
 & \frac{\eta_0' + k\eta_1'' + k^2\eta_2'' + \dots}{(1 + k\tau_1' + k^2\tau_2' + \dots)^2} - \frac{(\eta_0' + k\eta_1' + k^2\eta_2' + \dots)(k\tau_1'' + k^2\tau_2'' + \dots)}{(1 + k\tau_1' + k^2\tau_2' + \dots)^3} \\
 & + \frac{\xi_0' + k\xi_1' + k^2\xi_2' + \dots}{1 + k\tau_1' + k^2\tau_2' + \dots} = 0 \quad (11)
 \end{aligned}$$

where single and double primes denote respectively the first and second derivative with respect to γ .

Equating equal powers of k , we obtain

$$\begin{aligned}
 \eta_0'' - \eta_0' &= -a \sin k(\lambda\gamma + \alpha) \\
 \eta_0'' + \xi_0' &= 0 \quad (12)
 \end{aligned}$$

$$\begin{aligned}
 \eta_1'' - \eta_1' &= -\xi_0 a \cos k(\lambda\gamma + \alpha) + \xi_0' \tau_1'' + 2\xi_0'' \tau_1' - \eta_0' \tau_1' \\
 \eta_1'' + \xi_1' &= \eta_0' \tau_1'' + \xi_0' \tau_1' + 2\eta_0'' \tau_1' \quad (13)
 \end{aligned}$$

$$\begin{aligned}
 \eta_2'' - \eta_2' &= \frac{1}{2} a \xi_0^2 \sin k(\lambda\gamma + \alpha) - a(\xi_1 + \lambda\tau_1) \cos k(\lambda\gamma + \alpha) \\
 & + \xi_0' \tau_2'' + \xi_1' \tau_1'' + 2(\xi_0'' \tau_2' + \xi_1'' \tau_1') \\
 & - \eta_0' \tau_2' - \eta_1' \tau_1' - 3\xi_0' \tau_1' \tau_1'' + \xi_0'' \tau_1'^2 \\
 \eta_2'' + \xi_2' &= \xi_0' \tau_2' + \xi_1' \tau_1' + \eta_0' \tau_2'' + \eta_1' \tau_1'' \\
 & + 2(\eta_0'' \tau_2' + \eta_1'' \tau_1') - 3\eta_0' \tau_1' \tau_1'' + \eta_0'' \tau_1'^2 \quad (14)
 \end{aligned}$$

The converted initial conditions are

$$\begin{aligned}
 \eta_0(0) &= \eta_0'(0) = 0 \quad \xi_0'(0) = (u_0 - U)/f \\
 \eta_0'(0) &= v_0/f \quad (12a)
 \end{aligned}$$

$$\eta_1(0) = \eta_1'(0) = \xi_1'(0) = \eta_1''(0) = 0 \quad (13a)$$

$$\eta_2(0) = \eta_2'(0) = \xi_2'(0) = \eta_2''(0) = 0 \quad (14a)$$

Clearly the solutions of systems (12), (13), (14), ..., give respectively the zeroth-order, first-order, second-order, ..., solution of the nonlinear differential equations (5) and (6). It may be noted that system (12) is made equivalent to the linearized system [Kao and Neiburger, 1959] by letting $x(t) = Ut + x(0)$ in the right-hand side of (1). The problem is then to solve

for the successive systems (12), (13), (14), \dots , where
with the respective initial conditions (12a),
(13a), (14a), \dots .

Let us introduce a complex variable

$$\zeta = \xi + i\eta \quad (15)$$

where

$$\zeta = \zeta_0 + k\zeta_1 + k^2\zeta_2 + \dots \quad (16)$$

and

$$\zeta_j = \xi_j + i\eta_j \quad j = 0, 1, 2, 3, \dots$$

Multiplying the second equation of systems (12), (13), (14), \dots , by imaginary number i and combining with the first equations of the respective systems, we have

$$\zeta_0'' + i\zeta_0' = -a \sin k(\lambda\gamma + \alpha) \quad (17)$$

$$\zeta_1'' + i\zeta_1' = -a\xi_0 \cos k(\lambda\gamma + \alpha) + \xi_0'\tau_1'' + 2\xi_0''\tau_1' - \eta_0'\tau_1' \quad (18)$$

$$+ i\{\eta_0'\tau_1'' + (\xi_0' + 2\eta_1'')\tau_1'\} \quad (18)$$

$$\zeta_2'' + i\zeta_2' = \frac{1}{2}a\xi_0^2 \sin k(\lambda\gamma + \alpha) - a(\xi_1 + \lambda\tau_1) \cos k(\lambda\gamma + \alpha) \quad (19)$$

$$+ \xi_0'\tau_2'' + \xi_1'\tau_1'' + 2(\xi_0''\tau_2' + \xi_1''\tau_1') - \eta_0'\tau_1' - \eta_1'\tau_1' - 3\xi_0'\tau_1'\tau_1'' + \xi_0''\tau_1'^2 \quad (19)$$

$$+ i\{\xi_0'\tau_2' + \xi_1'\tau_1' + \eta_0'\tau_2'' + \eta_1'\tau_1'' + 2(\eta_0''\tau_2' + \eta_1''\tau_1') - 3\eta_0'\tau_1'\tau_1'' + \eta_0''\tau_1'^2\} \quad (19)$$

The converted initial conditions become

$$\zeta_0(0) = 0 \quad \zeta_0'(0) = \frac{1}{f} \{(u_0 - U) + iw_0\} \quad (17a)$$

$$\zeta_1(0) = \zeta_1'(0) = 0 \quad (18a)$$

$$\zeta_2(0) = \zeta_2'(0) = 0 \quad (19a)$$

.....

The calculations of the higher order of approximation are generally long and tedious. Since we are concerned with the case of small values of k , we shall carry the calculation as far as the first order of approximation.

For the convenience of examining the oscillations and trajectories of the particle in the physical plane, we shall express the solution in terms of the complex variable

$$Z(t) = x(t) + iy(t) \quad (20)$$

$$Z(t) = Z_0(t) + kZ_1(t) + k^2Z_2(t) + \dots$$

and

$$Z_j(t) = x_j(t) + iy_j(t) \quad j = 0, 1, 2, \dots \quad (21)$$

$Z_0(t)$, $Z_1(t)$, $Z_2(t)$, \dots , are respectively the zeroth-order, first-order, second-order \dots position vector of the particle. Time derivative of these quantities, $Z_0'(t)$, $Z_1'(t)$, $Z_2'(t)$, \dots , are respectively the zeroth-order, first-order, second-order \dots , velocity of the particle.

There are generally two types of solutions depending on whether $k\lambda$ is an integer or not. We shall discuss these solutions separately.

The case of $k\lambda$ not being an integer. In this

case the zeroth-order velocity of the particle which satisfies (17) and the initial condition (17a) takes the form

$$\begin{aligned} Z_0'(t) = & U - \frac{A}{1 - (k\lambda)^2} \\ & \cdot \{k\lambda \cos k(U_\tau t + \alpha) \\ & - i \sin k(U_\tau t + \alpha)\} \\ & + \left\{ Z_0'(0) - U + \frac{A}{1 - (k\lambda)^2} \right. \\ & \left. \cdot (k\lambda \cos k\alpha - i \sin k\alpha) \right\} e^{-ist} \quad (22) \end{aligned}$$

The first term of the right-hand side of (22) gives the effect of the mean pressure force of the system. The second term, which gives the effect of the sinusoidal part of the pressure field, may

represented by a vector describing an ellipse with an angular velocity equal to kU_r , as shown schematically in the hodograph (Fig. 1). The α term in (22), which gives the effect of the rotation of the system, may be represented by a vector describing a circle, in the opposite sense of the rotation of the system, with the orbital period. Combination of these three terms gives the resultant hodograph, as shown by the solid curve in Figure 1. It is to be noted that the amplitude of the zeroth-order oscillation depends

not only on the amplitude of the pressure wave but also on the magnitude of $(k\lambda)_2 - 1$. We see from (22) that the closer the magnitude of $(k\lambda)_2$ is to 1 the larger is the amplitude of the zeroth-order velocity of the particle.

It can be shown that the system of equations (5) and (6) contains no singularities in its first-order differential equations. We may therefore put $\tau_1 = 0$ in (13) and (18). If we impose the initial conditions (18a), the first-order solution of (18) is

$$\begin{aligned} v'(t) = & \frac{Av_0}{f[1 - (k\lambda)^2]} \{k\lambda \sin k(U_r t + \alpha) + i \cos k(U_r t + \alpha)\} \\ & + \frac{A^2}{2f[1 - (k\lambda)^2][1 - (2k\lambda)^2]} \{2k\lambda \cos 2k(U_r t + \alpha) - i \sin 2k(U_r t + \alpha)\} \\ & + \frac{A(P^2 + Q^2)^{1/2}}{fk\lambda(2 + k\lambda)} \left\{ (1 + k\lambda) \sin \left[f(1 + k\lambda)t + k\alpha - \cos^{-1} \frac{P}{\sqrt{P^2 + Q^2}} \right] \right. \\ & - i \cos \left[f(1 + k\lambda)t + k\alpha + \cos^{-1} \frac{P}{\sqrt{P^2 + Q^2}} \right] \} \\ & - \frac{A(P^2 + Q^2)^{1/2}}{fk\lambda(2 - k\lambda)} \left\{ (1 - k\lambda) \sin \left[f(1 - k\lambda)t - k\alpha - \cos^{-1} \frac{P}{\sqrt{P^2 + Q^2}} \right] \right. \\ & + i \cos \left[f(1 - k\lambda)t - k\alpha - \cos^{-1} \frac{P}{\sqrt{P^2 + Q^2}} \right] \} \\ & - e^{-i\tau_1 t} \left\{ \frac{Av_0}{f[1 - (k\lambda)^2]} (k\lambda \sin k\alpha + i \cos k\alpha) \right. \\ & + \frac{A^2}{2f[1 - (k\lambda)^2][1 - (2k\lambda)^2]} (2k\lambda \cos 2k\alpha - i \sin 2k\alpha) \\ & + \frac{A}{fk\lambda(2 + k\lambda)} [-(1 + k\lambda)(Q \cos k\alpha - P \sin k\alpha) + i(Q \sin k\alpha + P \cos k\alpha)] \\ & \left. + \frac{A}{fk\lambda(2 - k\lambda)} [(1 - k\lambda)(P \sin k\alpha + Q \cos k\alpha) + i(Q \sin k\alpha - P \cos k\alpha)] \right\} \quad (23) \end{aligned}$$

where

$$\begin{aligned} &= -\frac{1}{2} \left\{ v_0 + \frac{A}{1 - (k\lambda)^2} \sin k\alpha \right\} \\ &= -\frac{1}{2} \left\{ u_0 - U + \frac{Ak\lambda}{1 - (k\lambda)^2} \cos k\alpha \right\} \quad (24) \end{aligned}$$

The first-order velocity contributes, therefore, in addition to the two frequencies which are the same as that of the zeroth-order oscillation,

three new frequencies, namely, $2kU_r$, $(f + kU_r)$, and $(f - kU_r)$. It is to be noted that the amplitude of the first-order oscillation of frequencies, $2kU_r$ and $(f - kU_r)$, becomes large when the value $k\lambda$ is near $\frac{1}{2}$, 1, or 2. It can be shown that $Z_2'(t)$ is of the same order of magnitude as $A^2 f^{-2}$. Thus, $Z'(t) = Z_0'(t) + kZ_1'(t) + O(A^2 k^2 f^{-2})$.

The contribution of the zeroth- and first-order approximation to the trajectory of the particle is $Z_0(t) + kZ_1(t)$, in which $Z_0(t)$ and

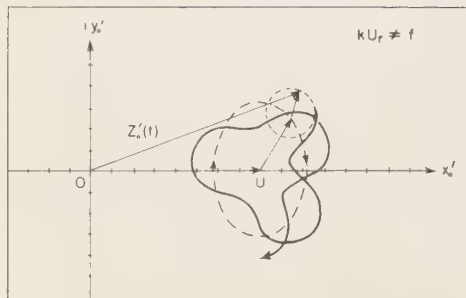


Fig. 1. Hodograph for the zeroth-order solution for the case of $k\lambda \neq 1$.

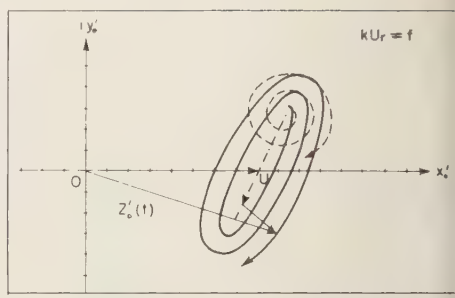


Fig. 2. Hodograph for the zeroth-order solution for the case of $k\lambda \neq 1$.

$Z_1(t)$ may easily be obtained by integrating (22) and (23), respectively. It is obvious that the trajectory, $Z_0(t) + kZ_1(t)$, consists of frequencies which are the same as those of the zeroth- and first-order velocities.

The case of $k\lambda = 1$. In this case the zeroth-order velocity of the particle which satisfies (17) and the initial conditions (17a) takes the form

$$Z_0'(t) = U + \frac{1}{2}Ae^{ik\alpha} \cos ft + \{Z_0'(0) - U - \frac{1}{2}Ae^{ik\alpha} - \frac{1}{2}Ae^{-ik\alpha}t\}e^{-ift} \quad (25)$$

The oscillation of the particle is no longer periodic. The amplitude of the zeroth-order velocity increases linearly with time. This is

a resonance phenomenon. The hodograph of the zeroth-order velocity of the particle consists of three characteristic parts, as shown schematically in Figure 2: (a) A constant translatory velocity U . (b) A simple harmonic oscillation along a straight line with the inertial period and amplitude $\frac{1}{2}A$. (c) A velocity, the amplitude of which grows linearly with time, spinning outward in the sense opposite that of the rotation of the system. A combination of these three terms gives the resultant hodograph, as shown by the solid curve in Figure 2, in which case a counterclockwise rotation of the system is assumed.

If we impose the initial conditions (18a), the first-order solution of (18) for $k\lambda = 1$ is

$$\begin{aligned} Z_1'(t) = & \frac{A^2}{12} t \{-2 \sin 2(ft + k\alpha) + i[3 - \cos 2(ft + k\alpha)]\} \\ & - \frac{A}{2} v_0 t \{\cos (ft + k\alpha) - i \sin (ft + k\alpha)\} - \frac{A}{2} \frac{v_0}{f} e^{ik\alpha} \sin ft \\ & + \frac{A}{36f} \left\{ \left[-5A \cos 2(ft + k\alpha) + 12 \left(u_0 - U - \frac{A}{2} \cos k\alpha \right) \cos (2ft + k\alpha) \right. \right. \\ & \left. \left. + 12v_0 \sin (2ft + k\alpha) \right] + 2i \left[2A \sin 2(ft + k\alpha) \right. \right. \\ & \left. \left. - 3 \left(u_0 - U - \frac{A}{2} \cos k\alpha \right) \sin (2ft + k\alpha) + 3v_0 \cos (2ft + k\alpha) \right] \right\} \\ & + \frac{A}{36f} e^{-ift} \left\{ \left[9A + 5A \cos 2k\alpha - 12 \left(u_0 - U - \frac{A}{2} \cos k\alpha \right) \cos k\alpha - 12v_0 \sin k\alpha \right] \right. \\ & \left. + 4i \left[-A \sin 2k\alpha + 3 \left(u_0 - U - \frac{A}{2} \cos k\alpha \right) \sin k\alpha + 3v_0 \cos k\alpha \right] \right\} \\ & + \frac{A}{4f} \left\{ -A + 2i \left[\left(u_0 - U - \frac{A}{2} \cos k\alpha \right) \sin k\alpha + v_0 \cos k\alpha \right] \right\} \end{aligned} \quad (26)$$

we see from the above equation that the first-order velocity contributes, in addition to the inertial oscillation, (i) a high-frequency oscillation of twice the natural frequency and (ii) a velocity of which the amplitude grows linearly with time.

The contribution of the zeroth- and first-order approximation to the trajectory of the particle is $Z_0(t) + kZ_1(t)$, in which $Z_0(t)$ and $Z_1(t)$ may be obtained by integrating (25) and (26), respectively. It can be shown that the zeroth-order trajectory grows linearly with time. However, the first-order trajectory consists of a component which grows at the second power of time. Also it can be shown that the higher order of approximations give a higher power of growth rate. The approximate solution holds, therefore, only for a certain time interval within which the higher-order terms are negligible as compared with the approximate solution.

Application of the nonlinear solution. In recent years the Naval Research Laboratory has conducted a series of experiments in tracking the trajectories of constant-pressure balloons at the 300-mb level by a direction-finding network [Eastenbrook and Anderson, 1953]. Angell [1958] found that the variance of the zonal and meridional wind components for these balloon flights shows a peak at periods near 50 hours

and appreciable amplitude at periods near 12, 25, and 30 hours. However, the amplitude of inertial oscillation was found to be small. The oscillation, which has periods near 50 hours and which associates with long waves in the atmosphere, may well be explained by the zeroth-order solution (22) or the equivalent solution of the linearized system of equations [Kao and Neiburger, 1959]. Equation (22) also shows that the amplitude of the inertial oscillation depends primarily on the geostrophic deviation because the sample average of the variance of the velocity components reduces the geostrophic deviation and, therefore, the amplitude of the inertial oscillation. Since the zeroth-order solution consists only of the forcing and the inertial frequencies, the oscillations at periods near 12, 25, and 30 hours cannot be explained by the linearized theory.

We see from (23) that the first-order velocity gives three new frequencies, namely, $(f + kU_r)$, $(f - kU_r)$, and $2kU_r$. In the middle latitudes, where the constant-pressure balloon flights were made, $f \sim 10^{-4} \text{ sec}^{-1}$ and the inertial period $T_f \sim 18$ hours. We conclude from the analysis of the zeroth-order solution that the forcing frequency associated with long waves, kU_r , is equal to $2\pi/50 \text{ hr} \sim 3.5 \times 10^{-5} \text{ sec}^{-1}$. The new frequencies and the corresponding periods may be calculated as follows:

$f + kU_r$	$2kU_r$	$f - kU_r$	T_{f+kU_r}	T_{2kU_r}	T_{f-kU_r}
$(10^{-4} \text{ sec}^{-1})$	$(10^{-4} \text{ sec}^{-1})$	$(10^{-4} \text{ sec}^{-1})$	(hr)	(hr)	(hr)
1.35	0.70	0.65	13.3	25.0	27.7

The first-order solution shows that oscillations occur at the periods near 13.3, 25, and 27.7 hours, which agree reasonably well with 12-, 25-, and 30-hour periods calculated from the trajectories of 300-mb constant-pressure balloons.

Conclusions. An analysis of the nonlinear characteristics of the oscillation of a particle in a long wave of finite amplitude in a rotating fluid shows that the motion of the particle generally depends on the interaction of the inertial oscillation and the forcing oscillation of the long wave. The motion of the particle is oscillatory if the forcing frequency kU_r is not equal to the multiples of the natural frequency.

In this case, the zeroth-order solution consists of the natural and forcing frequencies, whereas the first-order solution gives three new frequencies, namely, the sum and difference of the natural and forcing frequencies and a frequency of twice the forcing frequency. The motion of the particle is nonoscillatory if the forcing frequency is equal to the multiples of the natural frequency. In this case, the zeroth-order solution consists of the natural frequency with linearly growing amplitude, whereas the first-order solution gives a high-frequency oscillation of twice the natural frequency with linearly growing amplitude. A companion of the frequencies

predicted by the theory with those calculated from the trajectories of 300-mb constant-pressure balloons gives reasonably good agreement.

REFERENCES

- Andronow, A. A., and C. E. Chaikin, *Theory of Oscillations*, Princeton University Press, 1949.
- Angell, J. K., Lagrangian wind fluctuation at 300 mb derived from transosonde data, *J. Meteorol.*, **15**, 522-530, 1958.
- Forsythe, G. E., Exact particle trajectories for nonviscous flow in a plane with a constant Coriolis parameter, *J. Meteorol.*, **6**, 337-346, 1949.
- Kao, S.-K., and M. Neiburger, Oscillations and trajectories of particles in some pressure systems, *J. Geophys. Research*, **64**, 1283-1291, 1959.
- Kao, S.-K., and M. G. Wurtele, The motion of a parcel in a constant geostrophic wind field of parabolic profile, *J. Geophys. Research*, **64**, 765-777, 1959.
- Kryloff, N., and N. Bogolinboff, *Introduction Nonlinear Mechanics*, Princeton University Press, 1943.
- Lefschetz, S., *Lectures on Differential Equations*, Princeton University Press, 1946.
- Lighthill, M. J., A technique for rendering approximate solutions to physical problems uniformly valid, *Phil. Mag.*, **40**, 1179, 1949.
- Mastenbrook, H. J., and A. D. Anderson, Evaluation of transosonde system, *Mem. Rept. 2*, Naval Research Lab., Washington, 99 pp., 1950.
- Minosky, N., *Introduction to Non-linear Mechanics*, Edwards Bros., Ann Arbor, Mich., 1946.
- Poincaré, H., *Les méthodes nouvelles de la mécanique céleste*, v. 1, ch. IV, Gauthiers-Villars, Paris, 1892.
- Stoker, J. J., *Non-linear Vibrations*, Interscience Publishers, New York, 1950.
- Van der Pol, B., On relaxation oscillation, *Phil. Mag.*, **2**, 978-992, 1926.

(Manuscript received September 18, 1959.)

A Simplified Method for the Analysis and Synthesis of Dispersed Wave Trains

JAMES N. BRUNE, JOHN E. NAFE, AND JACK E. OLIVER

*Lamont Geological Observatory
(Columbia University)
Palisades, New York*

Abstract. A disturbance at one point of a dispersive medium resulting from an impulse applied at another point may be represented as a superposition of traveling plane waves. The phase and period of the disturbance at any instant are related by the principle of stationary phase to the phase and period of a traveling wave component. For the instantaneous phase of that traveling wave component the following equation may be written.

$$Ct - x = (N - \varphi_0/2\pi)CT$$

where C is the phase velocity, x the distance, T the period, t the travel time, N an integer, and φ_0 the initial phase of the traveling wave component. Since t and T may be measured from a record of the disturbance and x may be determined, the equation may be used to compute the phase velocity as a function of period, if the initial phases are known. If distance and the dispersion are known, initial phases may be determined. From distance, initial phases, and phase velocities the disturbance at any point may be constructed. The practical use of the method is demonstrated by application to antisymmetric waves in a cylindrical rod, Rayleigh waves from United States and Russian nuclear explosions, Rayleigh waves from the Hudson Bay earthquake of January 30, 1959, and Love waves from the Fairview Peak and Fallon, Nevada, earthquakes of 1954.

Analysis procedure. If a disturbance is applied at a given point in a dispersive medium, the motion at any other point is determined by the phase velocities of the spectral components and the characteristics of the initial disturbance. The method presented in this paper relates the phases of the various spectral components at the origin, their phase velocities, and the phases at the point of observation. If two of these quantities are given, the third may be determined in most cases. For example, the initial phases of the spectral components at the origin may be determined from a record of the motion at a distant point if the dispersive properties of the intervening medium are known. If the initial phases at the origin are known, the phase velocities in the intervening medium may usually be found from a record of the motion. If distance, initial phases, and phase velocities are known, the disturbance at any point as a function of time may be constructed. The application of the method is straightforward and can be under-

stood without involved mathematical treatment. The method is illustrated here by its application to the dispersion of elastic surface waves as encountered in seismology and to the dispersion of antisymmetric elastic waves in a long cylindrical rod.

The dispersed wave trains of seismograms are usually analyzed by measuring the period of vibration at each arrival time on the record and computing group velocity as a function of period. Group velocities thus obtained are compared with those for various models, and deductions are made about the structure along the propagation path. However, much valuable information is also contained in the actual phases of the recorded motion. Considerations of the phases have generally been limited to measurements of phase velocity by correlation of peaks between relatively closely spaced stations [Press, 1956; Ewing and Press, 1959]. In this paper the phase relationships of dispersed waves are considered in more detail.

We consider that type of wave propagation for which the disturbance, a function of space and time, may be represented by a superposition of traveling plane waves:

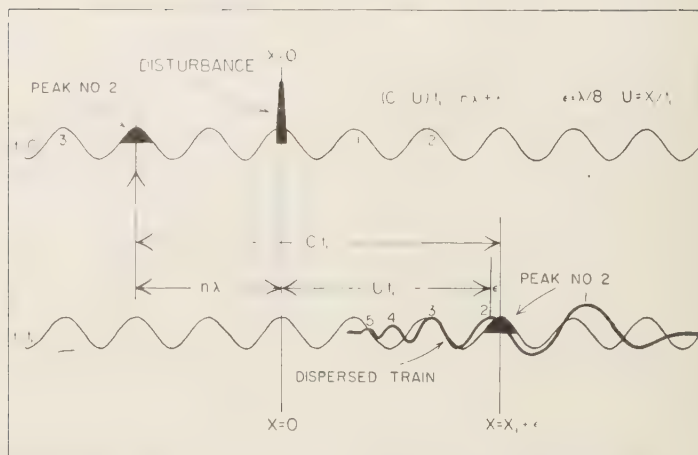


Fig. 1. Diagram indicating the conventions established for numbering peaks and illustrating the physical significance of equation (4). The diagram shows the spatial position of a traveling wave component at the initiation of a simple symmetric disturbance and at a later time when a dispersed wave train is arriving at the recording station. The initial phase of the traveling wave component φ_0 is zero. The dispersion is normal ($dU/dT > 0$) and consequently at $t = t_1$ the peak of the traveling wave component is $\lambda/8$ farther from the source than the corresponding peak of the total motion.

$$u(x, t) = \frac{1}{2\pi}$$

$$\cdot \int_0^\infty A(k) \cos(\omega t - kx + \varphi(k)) dk \quad (1)$$

where u is a component of the displacement, x a space coordinate, t the time, $A(k)$ an amplitude function, k the wave number, ω the angular frequency, and $\varphi(k)$ the initial phase as a function of wave number. This may easily be generalized to three dimensions if geometrical phase shifts are taken into account. Solutions of this type apply to many problems of propagation of elastic waves. The phase velocity $C = \omega/k$ is in general a function of period, so that (1) represents a dispersed wave train. We may visualize the disturbance as resulting from the interference of an infinite number of sinusoidal wave trains, each traveling with its own particular phase velocity.

Given the locations of all the component traveling waves at a certain time—at the initiation of the disturbance, for example—and given also the phase velocities with which they propagate, their positions may be determined at any later time. To find the resulting motion as a function of time at any distant point it is neces-

sary only to add by some means all the contributions from these traveling waves.

To trace the motion of the component traveling waves in space and time it is necessary to identify and distinguish points on each infinite sinusoidal wave train. In this paper we shall fix our attention on the peaks of the waves (any phase of the waves can be used) and identify various peaks by numbering them according to their distance from the origin at the time of the initiation of the disturbance (Fig. 1). A peak of 3.5 wavelengths behind the origin at $t = 0$ will be assigned the order number 3.5, and so on. A peak in front of the origin at $t = 0$ will be given a negative order number. In general, at $t = 0$, a peak of a traveling wave of frequency ω_0 will be located $N - \varphi(k_0)/2\pi$ wavelengths behind the origin, where N is an integer and $\varphi(k_0)$ is the initial phase of the wave. The quantity $N - \varphi(k_0)/2\pi$ is defined as the order number n of the peak. After a travel time t , the peak will arrive at a distance x from the origin. Considering Figure 1 and applying elementary principles, we may immediately write the equation for the motion of this peak:

$$Ct - x = (N - \varphi_0/2\pi)\lambda = n\lambda \quad (2)$$

where λ is the wavelength. In this equation

corresponds to the travel time of a peak of one of the component traveling waves. In general this may not correspond to the time t_1 of arrival of a peak of the total motion of the record, since the total motion is determined by the superposition of an infinite number of such traveling waves. However, when the stationary phase evaluation of (1) is applicable, a simple method may be used to determine the relation between the arrival time of peaks of the total motion and the arrival time of peaks of the component traveling waves. For large t —that is, when the wave train has become sufficiently dispersed—the integral (1) may be evaluated by the method of stationary phase and shown to be approximately

$$\sim \left\{ \frac{1}{2\pi t \left| \frac{d^2\omega}{dk^2} \right|_0} \right\}^{1/2} \cdot A(k_0) \cos [\omega_0 t - k_0 x + \varphi(k_0) \pm \pi/4] \quad (3)$$

provided ω_0 is not too near an extremum of group velocity and the amplitude does not vary too rapidly as a function of frequency [Pekeris, 1948]. In general these restrictions do not seriously limit the use of this equation. Peaks of the total motion will occur at a time t_1 and a distance x_1 when

$$\omega_0 t_1 - k_0 x_1 + \varphi(k_0) \pm \pi/4 = 2N\pi$$

or, in another form, when

$$\begin{aligned} \omega_0 t_1 - x_1 &= \left(N \mp 1/8 - \frac{\varphi(k_0)}{2\pi} \right) \lambda \\ &= (n \mp 1/8) \lambda \quad (\text{see Fig. 1}) \quad (4) \end{aligned}$$

A peak of dispersed motion with frequency ω_0 occurs when the peaks of the traveling wave components with frequencies equal to and near ω_0 constructively interfere with one another, but the phase of the total motion will be $\pi/4$ (or $T/8$, where T is the period of the wave) different from the phase of the component traveling wave of the same frequency.¹ The frequency of the total motion is determined by the customary method of measuring the slope of a plot of the peak number versus arrival time,

and the arrival time of the component traveling wave of this frequency is determined by adding or subtracting $T/8$ seconds from the arrival time of the peak of the total motion. For normal branches of dispersion ($dU/dT > 0$) the correction is subtracted from the arrival time, and for inverse dispersion ($dU/dT < 0$) the correction is added. Here U is the group velocity defined as x_1/t_1 [Ewing, Jardetzky, and Press, 1957, p. 144]. In problems where the method of stationary phase is not valid, the position of the peaks of the traveling wave components may be determined by Fourier analysis. Sato [1955, 1956] has described in detail the use of the Fourier transform to reconstruct the motion at the source and to derive phase velocity.

For the problem of cylindrical waves diverging from a source, the solution of the wave equation is the Hankel function of zero order of the second kind [Pekeris, 1948]. For distances greater than a few wavelengths, the asymptotic expansion of this Hankel function is valid.

$$H_0^{(2)}(kr) \sim (\sqrt{2/\pi kr}) \exp i(\pi/4 - kr) \quad (5)$$

Thus the phase of the total motion at a distance r from the origin will be

$$(\omega t - kr - \pi/4 \pm \pi/4)$$

The additional $-\pi/4$ phase shift is introduced into the total motion by the superposition of plane waves to form cylindrical waves. Thus for cylindrical wave patterns, a peak of the total motion arrives $T/8 \mp T/8$ seconds later than the peak of the plane wave of the same period, where the first $T/8$ results from the cylindrical nature of the problem and the second $T/8$ from the phase shift in the stationary-phase expansion. In practice these corrections can be made by simply subtracting 0 or $T/4$ sec from the travel time determined from the total motion as before. In the case of the spherical earth, solutions are in the form of Legendre polynomials but the asymptotic expansion given by Jahnke and Emde [1945, p. 117] shows that the $\pi/4$ correction in (5) also holds in this case.

The superposition of component traveling waves gives the dispersed wave train recorded at the station. The dispersed wave train has a certain number of peaks, each of which is associated with a peak of one of the component traveling waves. We may therefore assign the same order numbers to the peaks of the dispersed

¹ An additional $-T/8$ phase shift will be introduced for propagating cylindrical waves. The explanation for this follows later in the text.

TABLE 1. Description of Instruments

Instrument*	Constants†		Corresponding Figure and Amplification Curve in <i>Sutton and Oliver</i> [1959]	
15-80A	$\epsilon = 1.5\omega$	$\epsilon_g = 6\omega_g$	Fig. 2.	Ottawa vertical
15-80B	$\epsilon = 3\omega$	$\epsilon_g = \omega_g$	Fig. 3.	Resolute Bay horizontal
30-100	$\epsilon = 4.5\omega$	$\epsilon_g = \omega_g$	Fig. 5	
30-100P ₁	$\epsilon = 0.9\omega$	$\epsilon_g = 0.3\omega_g, \sigma\sigma_g = 4\epsilon\epsilon_g$	Fig. 5	
30-100P ₂	$\epsilon = 0.9\omega$	$\epsilon_g = 1.4\omega_g, \sigma\sigma_g = 4\epsilon\epsilon_g$	Fig. 5	

* Two numbers separated by a hyphen are used when referring to different types of instruments. The first number refers to the free period of the seismometer pendulum, the second to the free period of the galvanometer. In cases where instruments are considered which have the same set of free periods but which differ in other constants, a third term is added.

† σ = coupling coefficient of seismometer
 ϵ = damping coefficient of seismometer
 σ_g = coupling coefficient of galvanometer
 ϵ_g = damping coefficient of galvanometer

train as are assigned to the peaks of the component waves with which they are associated (Fig. 1). If the order numbers of the peaks of the various component waves are all integers, then the order numbers of the peaks of the dispersed train will all be integers. Equation (4) can be used to demonstrate the rate at which peaks develop on the record and therefore the order numbers to be expected. In practice a few trials will quickly establish the right order number. At the Airy phase it is convenient to define a time $t_0 = x/U_0$ for each distance x , where U_0 is the extreme value of the group velocity. The

period and the wavelength of motion at t_0 are constant with distance. If we consider λ to be constant and $t_0 = x/U_0$ to be the variable and solve equation (4) for n ,

$$n = \frac{(C_0 - U_0)}{\lambda} t_0 = \frac{(C_0 - U_0)}{\lambda} \frac{x}{U_0} \quad (6)$$

n will be the phase of the motion in wavelength at t_0 . The $\pm 1/8$ is canceled by $\mp 1/8$ from the other branch of motion, since the branches when added give the motion at the Airy phase [Pekeris 1948]. Therefore the phase at t_0 varies directly as the distance x and the difference between the

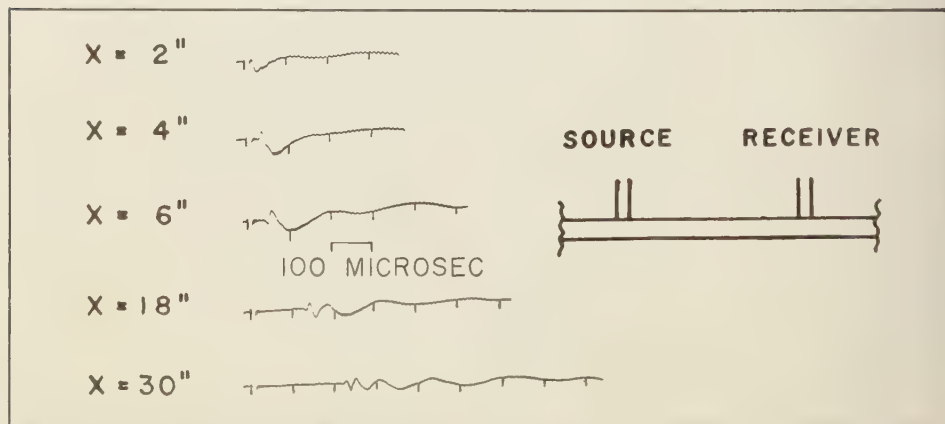


Fig. 2. Seismograms of the antisymmetric mode of vibration in a long cylindrical rod of hot rolled steel, 1 inch in diameter. Larger time marks are 100 μ sec apart. The origin time of the 10- μ sec pulse is at the first 100- μ sec time mark. The seismogram for $x = 2$ inches, at which distance the pulse has not been greatly dispersed, shows the direction of the initial motion.

ase and group velocities. Since the peaks of
e dispersed train emerge from (or merge into)
e Airy phase, (6) indicates how many peaks
l emerge in a given distance of travel. There-
re (6) indicates the order numbers to be
ected on the record and in particular the
ler number at t_0 . It shows that the numbering
peaks on the record is continuous for each
anch of the motion and the number of peaks
hich emerge from the Airy phase is the same
both the short-period and long-period
anches. In particular, if $\varphi(k) = 0$, that is, if
the waves were initially in phase as peaks at
e origin, then each of the peaks of the actual
motion would have an order number which is
integer. Although each peak changes period
it progresses from the Airy phase, it keeps
e same order number. If the order number
re suddenly to jump from one integer to
other, a sharp discontinuity in the phase
velocity would be implied and (3) would not be
id. Thus each peak on a given branch of the
motion has a different order number and the
secutive order numbers differ by an integer,
1, 1.5, 2.5, 3.5, etc. The direction of numbering
in the direction of increasing time on the

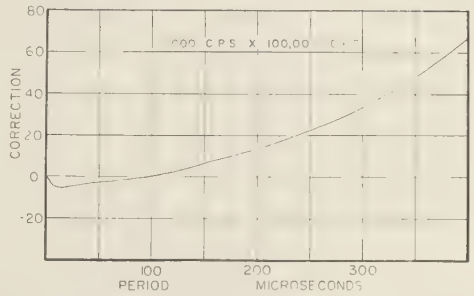


Fig. 3. Phase shift in microseconds for Kron-hite filter used for seismograms in Figure 2. Correction is applied to the travel time of the component traveling wave of a given period. The phase shift of the other apparatus is not significant for the periods used in this experiment.

record (Fig. 1). The order number of a peak of the dispersed wave train is one of its distinguishing properties. If a peak on the dispersed wave train has order number 2, that peak represents the constructive interference of those peaks of the traveling wave components that were located two of their own wavelengths behind the origin at the initiation of the disturbance.

If we have the time of arrival of the various

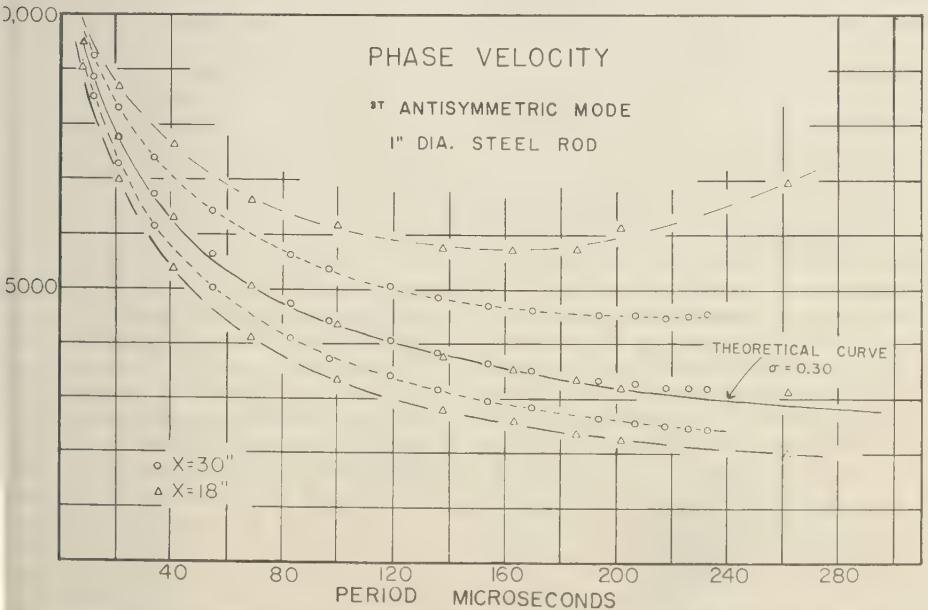


Fig. 4. Possible phase-velocity curves determined from seismograms given in Figure 2 for different orderings of peaks. The theoretical curve is taken from Hudson [1943].

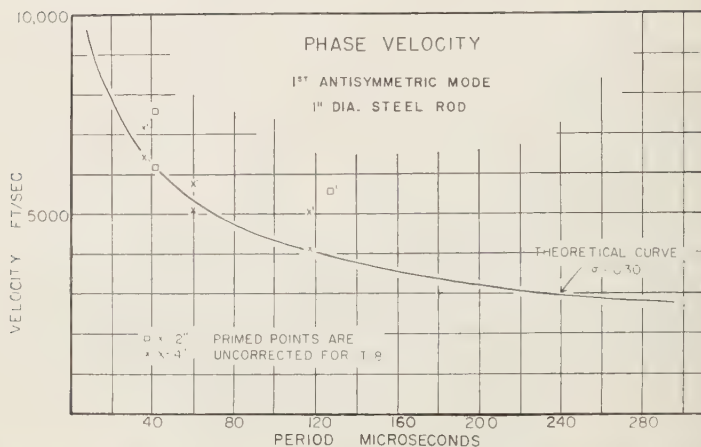


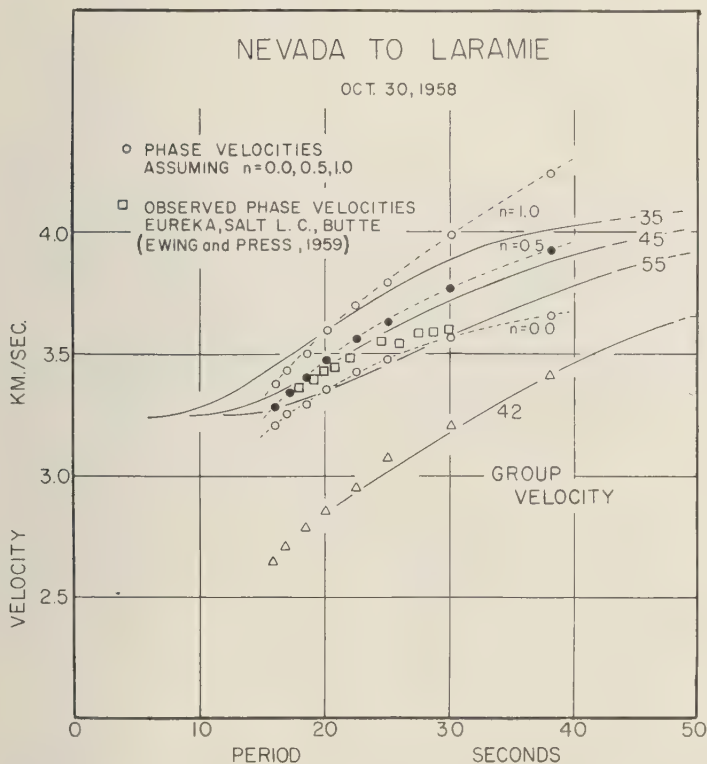
Fig. 5. Phase velocity determined from seismograms in Figure 2 for distances of 2 inches and 4 inches. The effect of the $T/8$ difference in travel time is indicated by the primed points.

peaks of the component waves on the record, it is necessary in practice to correct for the phase shift of the instrument. This is done by correcting the apparent arrival time of the peaks on the record. Phase-shift curves may be computed if the constants of the instrument are known. The phase-shift curves for a number of different instruments for different values of damping and coupling are given by Hagiwara [1958]. For the examples used in this paper, the phase-shift curves were computed for the instruments used at Palisades (Fig. 16). It was found, however, that the phase-shift curves of Hagiwara could be extrapolated with sufficient accuracy to be used without actual computation of the curve for the instruments used. Table 1 gives the constants for the instruments shown in Figures 17 and 18.

Antisymmetric waves in a long cylindrical rod. As an example of the use of this method for determining phase velocity, consider the case of waves of the first antisymmetric mode in a long (effectively infinite) cylindrical rod. Figure 2 shows seismograms at distances of 2, 4, 6, 18, and 30 inches. The source of the disturbance was a pulse of about 10 microseconds' duration transmitted to the rod by a barium titanate transducer. The detector was a barium titanate transducer located along the rod on the same side as the source (Fig. 2). The initial disturbance is symmetric with respect to distance along the rod, and from the direction of the first motion

recorded near the source it may be assumed that the waves were initially in phase as trough at the origin. Figure 4 shows possible phase velocity curves computed from (4) by assigning different order numbers to a certain trough on the record and numbering the other phases accordingly. The phase-shift correction for the filter in the receiving apparatus is given in Figure 3. Phase shifts introduced by the rest of the apparatus in the receiving circuit are relatively small. Because of the assumption of troughs at the origin, troughs were assigned integral order numbers and peaks were assigned half-integral order numbers. The theoretical curve shown in Figure 4 was taken from Hudson [1943] for the case of Poisson's ratio $\sigma = 0.30$. One of the curves determined from (4) gives excellent agreement with the theoretical curve, and the others may definitely be eliminated as having resulted from the wrong choice of order numbers. Thus, by a comparison of the experimental curve with the theoretical curves, it is possible to select the correct experimental curve from the infinity of discrete possibilities. The close agreement between this curve and the theoretical curve is additional justification for the assumption that troughs were in phase at the origin.

The determination of phase velocity from a single seismogram is dependent on the assumption of phases at the origin. In cases where we cannot assume the phases at the origin with confidence, it is still possible to determine the



6. Group velocity and possible phase velocities of Rayleigh waves recorded at Laramie, Wyoming, from nuclear explosion Blanca in Nevada.

se velocity if a number of recording stations available at different distances along a path, divided the dispersive characteristics do not γ . The spacing between the possible curves increases with increasing distance; thus, of all possible curves determined from a single station, only one will be in agreement with the possible curves determined at several other stations for the same initial phases. With a number of recording stations it is therefore possible to determine by elimination both the phase velocity and the initial phases. Figure 4 shows possible phase-velocity curves for distances of 18 and 30 inches on the model above, if heights are assumed at the origin. At both distances one of the possible curves agrees with the theoretical curve. It may be readily verified that this is the only reasonable curve compatible with records from both distances and with the requirement of fixed initial phases. Thus from

two stations both the phase-velocity curve and the initial phases have been determined. Although this method is similar to the conventional method of measuring the velocity by correlating the peaks between two stations, it has the advantages that it is not necessary to have the same peak appear on both records, that the possibility of misidentifying peaks is reduced, and that information about the source may be obtained.

If the phase-velocity curve for $\sigma = 0.30$ is assumed, the result from data from a single station is a straightforward determination of the phases at the origin. At distances of 2 and 4 inches from the origin, small phase differences have a large effect on the computed phase velocities, and hence small differences in the phases at the origin may be detected. In Figure 5 points are shown which result from the analysis of the seismograms in Figure 2 for distances of 2 and 4 inches (phases in addition to peaks and

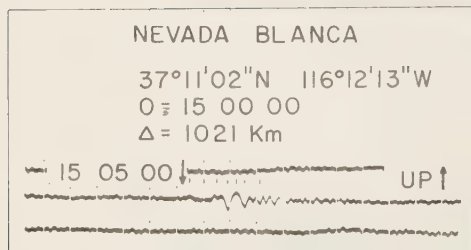


Fig. 7. Seismogram of Rayleigh waves recorded at Laramie, Wyoming, Oct. 18, 1958, from nuclear explosion Blanca in Nevada. The instrument is designated 30-100 P1 in Figures 17 and 18. Times and locations were released by the Atomic Energy Commission. Time marks are shown every half minute. The arrow indicates corrected time. The reading correction is -7 sec.

troughs were considered). If troughs are assumed at the origin, the unprimed points are obtained and they fall very nearly on the theoretical curve. The zero-order trough is shown clearly on the seismogram for 2 inches. If the $\pi/4$ correction had not been taken into account, the primed points would have resulted. This may be regarded as an experimental verification of the $\pi/4$ in the stationary phase approximation. The great sensitivity of this method at these distances occurs because the wavelengths involved are of the order of the distance.

Long-period Rayleigh waves from nuclear explosions. In demonstrating the application of this method to surface waves of the earth, cases are first considered for which the initial phases are fairly certain. Nuclear explosions, in particular, are localized both in time and space and for the longer seismic periods may be assumed to be nearly delta functions or spike functions with very slowly varying frequency spectra. Figure 6 shows the results of analysis of the long-period Rayleigh wave data recorded at Laramie, Wyoming, from the underground nuclear explosion Blanca in Nevada. The record is shown in Figure 7 along with pertinent data. Times (GCT throughout this paper) and locations are those given by the Atomic Energy Commission. Three possible phase-velocity curves are shown in Figure 6. The upper curve was computed from (4) on the assumption that the first peak on the record (uncorrected arrival time, 15 05 14) has order number 1, as indicated by writing $n = 1$ next to the curve in Figure 6. This would

mean that the traveling plane waves were initially in phase as peaks at the origin. The middle curve is based on the assumption that this same peak has peak order number 0.5, corresponding to traveling waves initially in phase as troughs at the origin. Similarly, the lower curve is based on the assumption that this peak has order number 0.0, corresponding to peaks at the origin. Phase shift due to the instrument is taken into account in all cases. The Rayleigh wave phase-velocity curves [Press, 1956] for crustal thicknesses of 35, 45, and 55 km are superimposed for comparison along with points observed by Ewing and Press for the Nevada-Utah-Wyoming region. The phase-velocity data indicated by the solid circles in Figure 6 are in reasonable agreement with the data of Ewing and Press [1959] and give a crustal thickness of about 42 km, which is in good agreement with that deduced from group-velocity data. The data indicate that the waves were initially in phase as troughs at the origin.

Since explosions on the earth's surface result in a downward compressive force on the surface, the traveling wave components which make up the Rayleigh wave train may be expected to be in phase as troughs at the origin. For ordinary Rayleigh wave motion, troughs of vertical displacement are associated with compressions of the medium in which the waves are propagating.

In order to check the phases to be expected at the origin from an explosion, a model experiment was performed in which a spark model of the explosion. With the spark source very near the dispersive medium (single surface layer on a two-dimensional disk model), analysis of the recorded motion at a distance of 10 inches definitely proved that the Rayleigh wave phases were troughs at the origin. With the spark at greater elevations (up to $1/2$ inch) from the surface of the model, the same result was obtained if the time of origin was taken to be the instant the shock wave from the spark reached the model. Additional studies of Rayleigh waves generated by nuclear explosions high in the atmosphere have been made by Pomeroy [1959], and his data confirm the nature of the source as described here. Thus troughs of vertical motion may be expected for the initial phases of Rayleigh waves from explosions in the air.

Very shallow underground explosions can

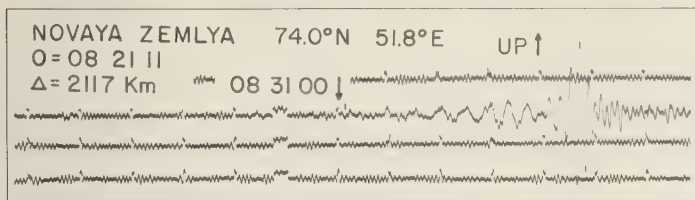


Fig. 8. Seismogram from the Russian nuclear explosion of October 22, 1958, recorded at Uppsala, Sweden. The instrument is designed 15-80 A in Figures 17 and 18. One-minute time marks are shown on the record. The arrow indicates corrected time. The reading correction is -4 sec.

umed to be effectively on the surface for the long wavelengths considered here. This assumption is supported by Rayleigh wave data recorded at Palisades from nuclear explosions in Nevada. The Rayleigh wave motion from the nuclear explosion Hood, which was fired on July 5, 1957, with the bomb suspended from a balloon, indicated exactly the same phases at the origin as that from the nuclear explosion Blanca, which was fired on October 30, 1958, with the bomb at a depth of about 1000 feet below ground.

The group-velocity data shown in Figure 6 indicate the same crustal thickness as the phase velocity. The group-velocity curve for 42 km (Fig. 6) was determined from the group-velocity data used by Press [1956] to construct by the usual method the phase-velocity curves used in the figure. The combined use of group velocity and phase velocity as presented in this paper offers a powerful method for the interpretation of crustal structure. If it may be assumed that a certain model of the earth is approximately

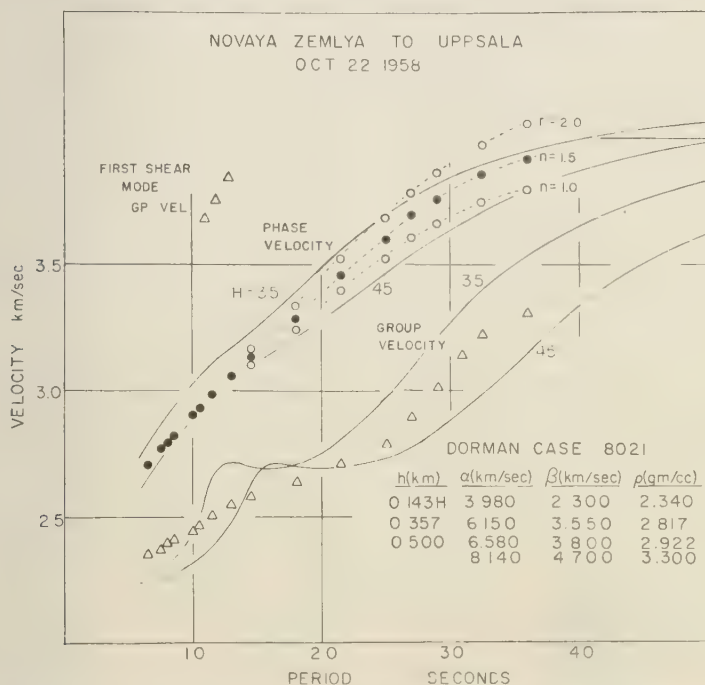


Fig. 9. Group velocity and possible phase velocities determined from the record shown in Figure 8. The center curve, labeled $n = 1.5$, was determined by assuming troughs of vertical motion at the origin.

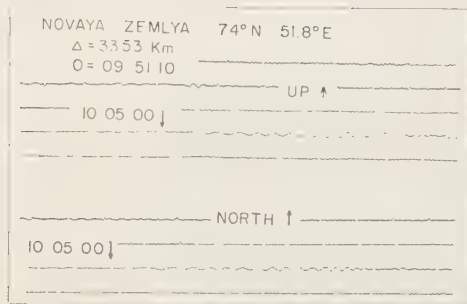


Fig. 10. Seismograms from the Russian nuclear explosion of October 18, 1958, recorded at Resolute Bay, N.W.T., Canada. The instrument is designated 15-80 B in Figures 17 and 18. One-minute time marks are shown on the record. The arrow indicates corrected time. The reading correction is ± 0 sec.

correct, then the group-velocity data may be used to fix the phase-velocity curve within narrow limits. A series of group- and phase-velocity curves may be computed for models differing only in the absolute values of crustal thickness, that is, models which have the same number of layers, the same ratios of layer thickness, and the same values of compressional-wave velocity, shear-wave velocity, and density for the respective layers, but which have different values for crustal thickness. Then only one of the series of possible phase-velocity curves will indicate the same crustal thickness as the group velocity indicates.

Rayleigh waves from Russian nuclear explosions of October 1958 were also studied by this method. Epicentral locations and origin times given in the bulletin of the seismograph station at Uppsala, Sweden, were used. These locations appear to be based on data from three Swedish stations only and so may not be as precise as that of a well-determined earthquake epicenter. However, the distance of the explosion of October 22, 1958, from Uppsala, Sweden, determined from this data is probably not greatly in error, since impulsive P-wave and impulsive S-wave arrivals were received at Kiruna, Sweden, which is on nearly the same azimuth from the explosion as Uppsala. Rayleigh waves from these explosions were recorded on long-period seismographs (15-80 A in Figures 17 and 18; see Table 1) operated at Uppsala. Records are shown in Figure 8. The group-

velocity data along with three possible phase velocity curves for different assumptions phases at the origin are shown in Figure 9. Both the phase-velocity and group-velocity data indicate that the average crust between Novaya Zemlya, U.S.S.R., and Uppsala has a layer of low-velocity material at the surface. The model used to compute the theoretical curve has a low-velocity layer 5 km thick at the surface (Fig. 9). With this model, the phase velocity curve that results from assuming troughs at the origin (middle curve) indicates approximately the same crustal thickness as the group-velocity data. The waves with period of about 13 sec arriving at 08 30 30 correspond to the M_2 , or first shear, mode. Excitation of waves of this mode varies from shot to shot throughout the series of nuclear explosions in the Soviet Union in October 1958. These waves have not been studied in detail for this paper.

The preceding example clearly shows the necessity of having a model that is approximately correct before attempting a precise interpretation of crustal structure by this method. If previous continental curves had been used, and if troughs had been assumed at the origin, the phase and group velocities would not have indicated consistent crustal thickness, and the crustal thickness indicated by the group-velocity data would be approximately 48 km in contrast with a crustal thickness of 40 km indicated by comparison of the same data with the model shown in Figure 9. This model was chosen to agree approximately with the group-velocity data, and the choice should not be considered a unique solution for the crustal structure.

Although it is necessary to have an accurate model before attempting precise interpretation of crustal thickness and initial phases by this method, in many cases the phases at the origin may be approximately determined without attempting to find a model that gives a good fit to the whole spectrum of data. This is due to the fact that for almost all paths on the earth's surface the phase velocity of Rayleigh waves is 4.0 ± 0.1 km/sec in the period range of 45 to 100 sec. If the distance from source to receiver is small enough, the difference in phase-velocity values at these periods, resulting from different assumptions of phases at the origin, will be so great that the actual phases at the origin may be specified within a narrow range. Of course

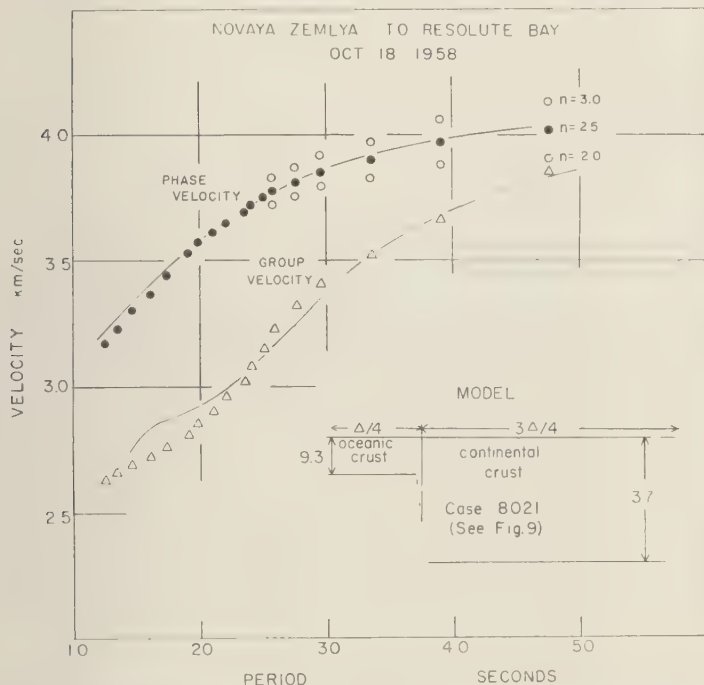


Fig. 11. Group velocity and possible phase velocities determined from the record shown in Figure 10. The center curve, labeled $n = 2.5$, was determined by assuming troughs of vertical motion at the origin.

struments of high magnification in the period range of 45 to 55 sec must be used, so that Rayleigh waves of these periods may be recorded. Long-period Rayleigh waves from Russian

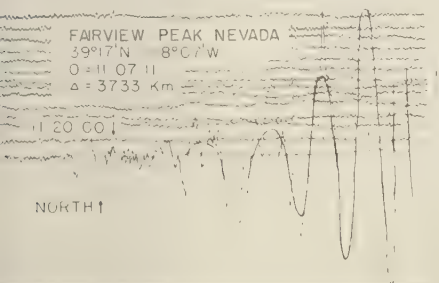


Fig. 12. Seismogram of the Fairview Peak earthquake recorded at Palisades Dec. 16, 1954. The phase shift curve for this instrument ($T_0 = 1$, $\tau = 75$) was taken from Hagiwara [1958]. One-minute time marks are shown on the record. The arrow indicates corrected time. The reading correction is +59 sec.

nuclear explosions were also recorded at Resolute Bay, N.W.T., Canada, on long period instruments similar to those at Uppsala (Fig. 10). For a reasonable assumption of the type of structure between Novaya Zemlya and Resolute Bay, the phase and group velocities determined from these data (Fig. 11) indicate that these Rayleigh waves were in phase as troughs at the origin. The structure was assumed to be continental over three fourths of the path and oceanic over the other fourth. Additional evidence would be required to determine the structure uniquely.

The application of this method is not limited to data from vertical-component instruments, of course. Ideally, the longitudinal horizontal motion of Rayleigh waves is $\pi/2$ out of phase with the vertical motion so that assigning to peaks of horizontal motion order numbers $N + 1/4, N + 3/4$, etc., is equivalent to assigning to corresponding peaks of the vertical motion integral or half-integral order numbers. This may readily be verified by computing various

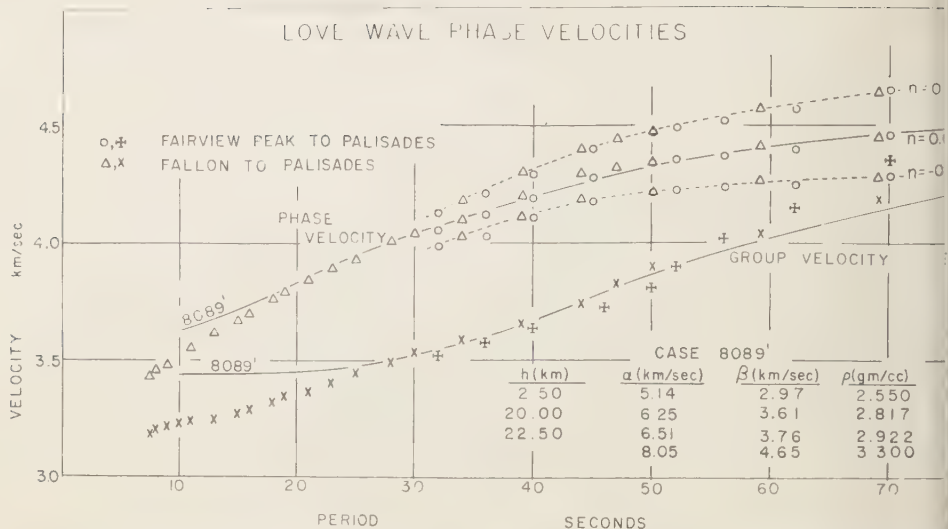


Fig. 13. Group velocity and possible phase velocities for Love waves from Nevada to Palisades, New York, determined from the seismogram in Figure 12 and the seismogram shown in *Ewing, Jardetzky and Press* [1957, p. 215]. $n = 0.0$ indicates that initial northward displacement was assumed at the origin.

possible phase velocities from the record shown in Figure 10 of the north-south component of ground motion resulting from the Russian nuclear explosion of October 18, 1958.

Continental Love waves. The method was also applied to Love waves recorded at Palisades

from the Fairview Peak and Fallon, Nevada earthquakes of 1954. The path is almost due east-west and is entirely continental. Seismograms are shown in Figure 12 and in *Ewing, Jardetzky, and Press* [1957, p. 215]. Three possible phase-velocity curves were computed from (

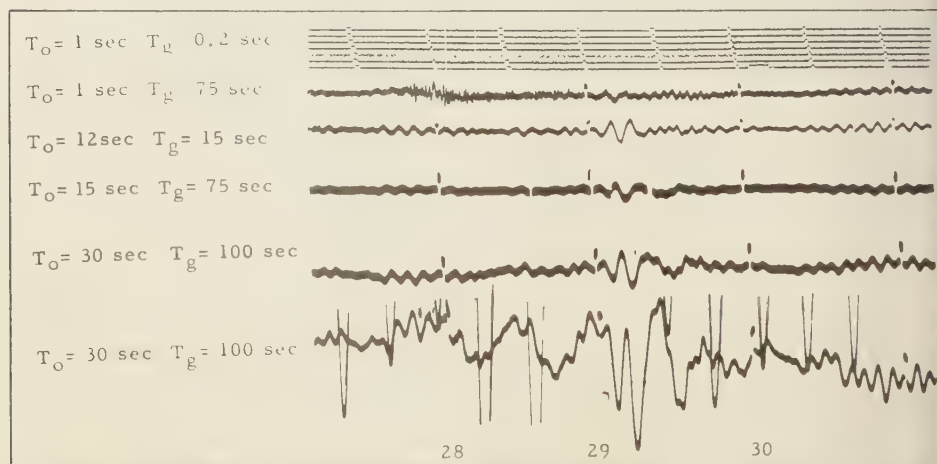


Fig. 14. Seismograms of the Hudson Bay earthquake of January 30, 1959, recorded at Palisades. The lower instrument is designed 30-100 P2 in Figures 17 and 18. One-minute time marks are shown on the records. The reading correction is +7 seconds.

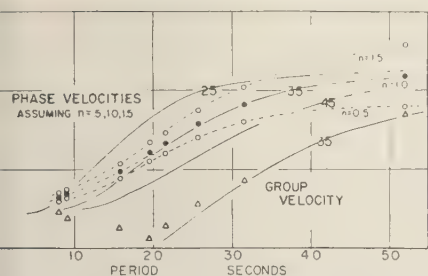


Fig. 15. Possible phase velocities and group velocity from Hudson Bay to Palisades determined from the record in Figure 14. $n = 1.0$ indicates that peaks of vertical motion were assumed the origin.

different assumptions of the order number of the first peak (Fig. 13). Phase-shift curves were taken from Hagiwara [1958]. The theoretical curve, Dorman case 8089, was calculated for a structure that has been found to be representative of continents. The data indicate that the waves were initially in phase as peaks at the origin, but, as northward displacement (accumulated strain) increased, the initial acceleration, of course, was southward when the strain was released. This would be expected because of the

known right-lateral motion of the faulting in this area. The right-lateral motion of the Fairview Peak fault is well determined from field evidence [Slemmons, 1957; Larson, 1957] and from precise surveys [Whitten, 1957]. A fault-plane solution for the Fairview Peak earthquake also shows right-lateral motion and a $N 11^\circ W$ strike of the fault [Romney, 1957]. Strike-slip motion on a fault is commonly thought to result from the release of accumulated strain produced by horizontal movements of one block of the earth's crust with respect to an adjacent block. In the case of the Nevada earthquakes, the right-lateral motion would result from a western block of crust moving northward and tending to drag the east block with it. Thus, the east block would have been initially strained to the north in a zone along the fault. The path to the recording station is nearly perpendicular to the strike of the fault. The nearly identical data from both the Fairview Peak and Fallon earthquakes are consistent with this interpretation, and indicate that right-lateral forces were also associated with the Fallon earthquake even though the surface expression of the fault did not show any lateral movement [Tocher, 1956].

Rayleigh waves from earthquakes. Rayleigh

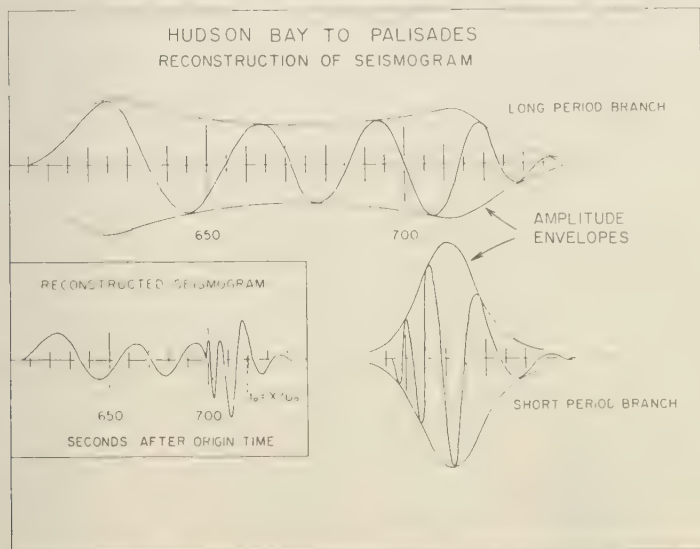


Fig. 16. Seismogram reconstructed from the group-velocity and phase-velocity data for $n = 1.0$, shown in Figure 15. This reconstructed seismogram is to be compared with the seismogram in Figure 14.

TABLE 2. Phase-Velocity Calculations, Hudson Bay to Palisades, January 30, 1959*

Corr. Arrival Time of Peak of Total Motion	Travel Time, sec	Period T , sec	$T/8$ $\pm T/8$, sec	Inst. Corr., sec	Corr. Travel Time t , sec	Order Number, n	nT	Phase Vel $X/(t - nT)$
05 27 56	624	52	-13	-10	601	1	52	4.09
28 34	662	31.5	-8	-16	638	2	63	3.91
29 03	691	25.7	-7	-13	671	3	77	3.78
29 29	717	21.5	-5	-12	700	4	86	3.66
29 49	737	19.5	-5	-11	721	5	97.5	3.61
29 09	697	8.0	-0	-5	692	2	16	3.325
29 17	705	9.0	-0	-6	699	3	27	3.335
29 29	717	15.7	-0	-9	708	4	62.5	3.48

* Distance = 2248 km; origin time = 05 17 32 GCT; reading correction + 7 sec.

Note: The group velocity is given by the distance divided by the group travel time indicated in column 8. These group travel times must be corrected for the effect of the instrument indicated in Figure 1.

waves from the Hudson Bay earthquake of January 30, 1959, were recorded especially well at Palisades over an unusually wide range of periods. Waves with periods from 1 to 53 sec were recorded on a variety of long- and short-period seismographs (Fig. 14). Only waves with periods greater than 9 sec were used here, however. Both branches of the dispersion curve in the vicinity of the minimum of group velocity at a period of about 19 sec were clearly discernible and the numbering of the peaks could be carried through the corresponding Airy phase without ambiguity. Three possible phase-velocity curves were computed from (4) for various assumptions of the order number of the first peak (Fig. 15). The initial phases for Rayleigh waves from earthquakes cannot be assumed as certainly as they can for explosions, and therefore it is not certain that the peaks on the record have order numbers that are exact integers or half-integers. But due to the local nature of most earthquakes and their short time duration, especially in the case of small shocks, it may be that this is often a reasonable assumption. One of the three curves fits Press's curves well and indicates a crustal thickness of approximately 35 km, which agrees with the thickness indicated by the group velocity. As before, the group-velocity curve corresponding to a crustal thickness of 35 km was determined from the group-velocity data of Press and others [1956]. This result for crustal thickness also agrees closely with the seismic refraction results that Hodgson [1953] obtained for the Canadian shield areas which are probably

similar in structure to the average path between Hudson Bay and Palisades. The mean thickness of the crust, based on P wave arrivals, was found by Hodgson to be 35 ± 5 km. Because of the wide separation of the possible phase-velocity curves at a period of 50 sec for order numbers differing by $1/2$, it is probable that the middle curve in Figure 15 is approximately correct and that initially the waves were in phase as predicted at the origin. The phase velocities for the short periods are higher than those indicated by Press's curves due to the nature of the path. Correction for any initial phases would have little effect on this part of the curve, since differences in the order numbers of $1/2$ shift the curve very little. The Airy phase in this example has a period of approximately 19.5 sec. The point on each side of the Airy phase match exactly the trend of the curve, and it is certain that the order number has been miscorrelated on one of the branches. If one of the branches had been numbered compatibly with the other branch, a discontinuity would have occurred in the resulting phase-velocity curve. Table 2 shows the calculations carried out in this example in order to illustrate the simplicity of the calculations required by this method. The corrections which are applied to the arrival times of the various phases as a result of instrument phase shift may be found in Figure 17.

If the distance of a disturbance is known and the phase velocities and phases at the origin are known, (2) and (4) may be used to construct the resulting motion at the recording station.

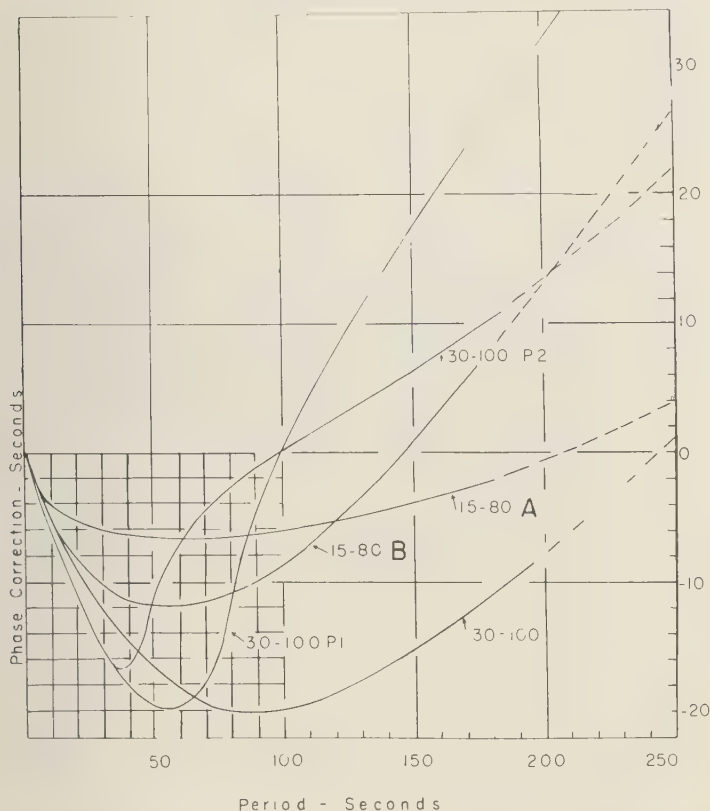


Fig. 17. Phase-shift correction in seconds for the long-period instruments used at Palisades. The first number for each instrument indicates the free period of the pendulum; the second number indicates the free period of the galvanometer. The instruments used in this paper were 15-80 A, 15-80 B, 30-100 P1 (underdamped) and 30-100 P2 (overdamped). See Table 1.

As a test of this method for reconstructing seismograms, (2) was used to compute the arrival time of phases on the seismogram (Fig. 6); the phase velocity for $n = 1.0$ from Figure 15 was used. The source was assumed to be a disturbance with all spectral components of equal amplitude and in phase as peaks of vertical motion at the origin. The computations were essentially the reverse of those indicated in Table 1. First, the arrival times of the various phases corresponding to various order numbers were computed from the phase velocity, the group velocity, and (2). Then the shift of the phases by the instrument was added. For the long-period branch the phases of total motion were constructed $T/4$ seconds later than the phases of the spectral components in order to

agree with the stationary phase approximation (3) and the geometrical phase shift (5). An amplitude envelope was computed for each branch using the coefficient in (3) for phases not near the Airy phase and using the Airy phase envelope given by Pekeris for phases near and in the Airy phase, in each case multiplying by the amplification of the instrument for the period of vibration arriving at each time on the record. The two branches were then added together to give the expected seismogram. Comparison of the reconstructed seismogram in Figure 16 with the actual seismogram recorded at Palisades in Figure 14 shows excellent agreement, both in phase and amplitude. From this it may be concluded that the assumption that the spectral components involved had equal

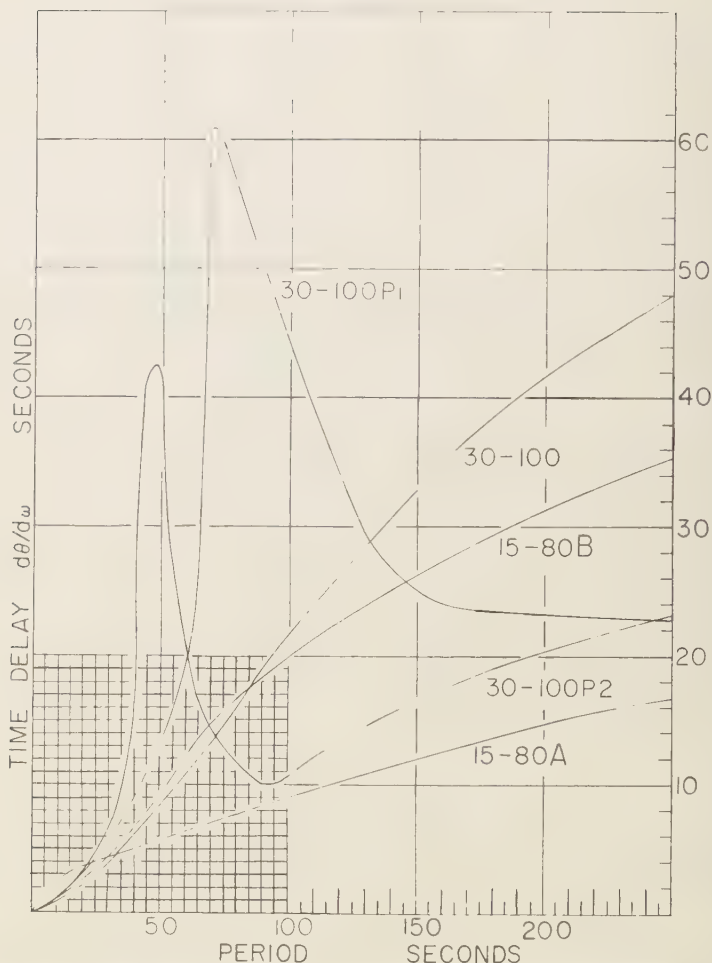


Fig. 18. Group time delay, $d\theta/d\omega$, for the long-period seismographs used at Palisades. In this figure is the phase shift of the instrument in radians. Also see Table 1.

amplitudes at the origin was very nearly correct (neglecting attenuation with distance in this restricted period range). The agreement of phase was, of course, assured beforehand, since the phases on the record were used to determine the phase velocity. It is of interest that the high relative amplitudes of waves corresponding to the long-period branch were mostly due to the peaked amplification of the instrument. For an instrument with constant amplification as a function of period, the short-period branch would have been several times larger in amplitude than the long-period branch.

The method presented in this paper offers many new possibilities for the advancement of seismology. Surface waves and dispersed waves of other types may be used to greater advantage. If a large number of long-period instruments were located at critical points around the world and surface wave data systematically analyzed by this method, the phase velocities over most of the earth could be determined in a relatively short period of time. Once phase velocities are well known for different paths over the earth's surface, the method can be used for detailed study of the nature of the motion at the source.

various types of seismic events. The dependence of initial phases on dip and strike of the fault may be established. Seismologists will then have a powerful tool for studying the direction of faulting in the earth's crust. The phase-velocity curves will also be important in the interpretation of the structure of the earth. In many cases the phase velocity may be used much more effectively than the group velocity for structural interpretation, and in practice the two sets of data complement each other, even though theoretically the phase-velocity curve completely determines the group-velocity curve.

Summary. The validity of a simple but powerful method for analyzing dispersed wave motion has been demonstrated by its application to several problems of elastic surface wave dispersion. The phenomenon of dispersion is easily visualized by this method. In many practical cases the method will lead to a unique determination of the phase-velocity curve for the path between the source and the station at which the motion is recorded or between two recording stations. The $\pi/4$ phase shift in the stationary-phase solution of dispersed wave motion has been verified experimentally. The initial phase of Rayleigh waves of long period from nuclear explosions both above and as much as about 1000 feet below the surface of the ground have been shown to be compressions (troughs of vertical motion). The phase-velocity curve for the path from Nevada to Laramie, Wyoming, has been determined and it indicates an average crustal thickness of about 42 km if the continental phase-velocity curves of Press are used. Phase-velocity curves and phases at the origin were also approximately determined for paths from Novaya Zemlya to Resolute Bay and Uppsala. The initial phases determined for Love waves from Nevada recorded at Palisades were found to be northward strain at the origin. The method has been used in reverse to reconstruct the motion on the seismogram for the Hudson Bay earthquake of January 30, 1959.

Acknowledgments. The authors wish to thank Maurice Ewing, George Sutton, James Dorman, Paul Pomeroy, and Maurice Major, staff members of the Lamont Geological Observatory, for valuable help and advice in carrying out this study. Thanks are due Markus Båth of the Seismological Laboratory, Uppsala, Sweden, for Uppsala Seismograms, and J. H. Hodgson of the Dominion Observatory of Ottawa, Canada, for the Resolute

Bay seismograms. Accumulation of the data under the IGY program was supported by National Science Foundation Grant Y/11.12/54. This work was also supported in part by the Office of Ordnance Research under Contract DA-30-069-ORD-2431, the Office of Naval Research under Contract N-onr 266 (53) and the Air Force Cambridge Research Center under Contract AF 19 (604) 4144.

REFERENCES

- Ewing, M., W. S. Jardetzky, and F. Press, *Elastic Waves in Layered Media*, McGraw-Hill, New York, 1957.
- Ewing, M., and F. Press, Determination of crustal structure from phase velocity of Rayleigh waves, Part III, the United States; *Bull. Geol. Soc. Am.*, 70, 229-244, 1959.
- Hagiwara, Takahiro, A note on the theory of the electromagnetic seismograph, *Bull. Earthquake Research Inst., Tokyo Univ.*, 36, 139-164, 1958.
- Hodgson, J. H., A seismic survey in the Canadian Shield, I, Refraction studies based on rockbursts at Kirkland Lake, Ont., *Dominion Observatory Ottawa Publ.*, 16, 113-163, 1953.
- Hudson, G. E., Dispersion of elastic waves in solid circular cylinders, *Phys. Rev.*, 63, 46-51, 1943.
- Jahnke, E., and F. Emde, *Tables of Functions*; Dover Publications, New York, 1945.
- Larson, E. R., Minor features of the Fairview Fault, Nevada, *Bull. Seis. Soc. Am.*, 47, 377-386, 1957.
- Oliver, Jack, and Maurice Ewing, Seismic surface waves at Palisades from explosion in Nevada and the Marshall Islands; *Proc. Natl. Acad. Sci.*, 44, 780-785, 1958.
- Pekeris, C. L., Theory of propagation of explosive sound in shallow water; *Geol. Soc. Am. Mem.*, 27, 1948.
- Pomeroy, Paul, J. Oliver, and M. Ewing, Seismic waves from high altitude nuclear explosions, in preparation.
- Press, Frank, Determination of crustal structure from phase velocity of Rayleigh waves, Part I: Southern California; *Bull. Geol. Soc. Am.*, 67, 1647-1658, 1956.
- Press, Frank, M. Ewing, and J. Oliver, Crustal structure and surface wave dispersion in Africa, *Bull. Seis. Soc. Am.*, 46, 97-103, 1956.
- Romney, Carl, Seismic waves from the Dixie Valley-Fairview Peak earthquakes, *Bull. Seis. Soc. Am.*, 47, 301-320, 1957.
- Sato, Yasuo, Analysis of dispersed surface waves by means of Fourier Transform I, *Bull. Earthquake Research Inst., Tokyo Univ.*, 33, 33-48, 1955.
- Sato, Yasuo, Analysis of dispersed surface waves by means of Fourier Transform II. Synthesis of movement near the origin, *Bull. Earthquake Research Inst., Tokyo Univ.*, 34, 9-18, 1956.
- Sato, Yasuo, Attenuation, dispersion and the wave guide of the G wave; *Bull. Seis. Soc. Am.*, 48, 231-251, 1958.

- Slemmons, David B., Geological effects of the Dixie Valley-Fairview Peak, Nevada, earthquakes of December 16, 1954, *Bull. Seis. Soc. Am.*, 47, 353-376, 1957.
- Sutton, George, and Jack Oliver, Seismographs of high amplification at long periods; *Ann. géophys.*, in press, 1959.
- Tocher, Don, Movement on the Rainbow Mountain fault, *Bull. Seis. Soc. Am.*, 45, 10-14, 1956.
- Whitten, C. A., Geodetic measurements in the Dixie Valley area; *Bull. Seis. Soc. Am.*, 47, 321-325, 1957.

(Manuscript received September 19, 1959.)

Aerial Gravity Measurements

LLOYD G. D. THOMPSON

*Geophysics Research Directorate
Air Force Cambridge Research Center
Bedford, Massachusetts*

AND

LUCIEN J. B. LACOSTE

*LaCoste and Romberg
Austin, Texas*

Abstract. Tests with an airborne LaCoste and Romberg Sea Gravity Meter (surface type) have shown that gravity measurements in a flying aircraft are feasible. At high altitudes the aircraft was a sufficiently stable platform for the gravity meter to operate satisfactorily. Analysis of in-flight acceleration problems has shown that gravity observations can be made at high aircraft velocities with proper flight programming and navigation systems. North, south, and west traverses over an Askania tracking range gave 5 min average gravity readings which plotted into smooth profiles. An accuracy of better than 10 mgal was obtained which meets requirements for geodetic applications where mean gravity values for $1^{\circ} \times 1^{\circ}$ squares are required.

Introduction. The problem of making gravity measurements in an aircraft has long been a challenge to geophysicists. While the solution has been most elusive, the results in this report show that such measurements are feasible for certain applications. Previous work has been done from the exploration point of view [for example, *Lundberg, 1957*] where very accurate and sensitive detecting devices are required and the problems seem to be insurmountable. If, however, the problem is approached from another standpoint where relatively inaccurate measurements are made (say 10 mgal), the problems are not nearly so formidable. The justification for such an approach lies in the applications or requirements for gravity measurements that are not too accurate, such as in regional or reconnaissance gravity surveys and for geodetic purposes. On this basis the Geodesy and Gravity Branch of Geophysics Research Directorate, Air Force Cambridge Research Center, initiated a program to develop an airborne gravity meter system to provide a means for rapid global coverage. It is of interest to note that the possibility of performing such gravity measurements and their usefulness has been

recognized even more recently by *Fedynskiy [1959]*.

The main problem areas to be considered for gravity measurements in an aircraft are (1) the type of observation that is required, (2) navigation including aircraft velocity, elevation, and space positioning, (3) in-flight accelerations of the aircraft (or the stability of the aircraft as a platform), (4) the Eotvos effect and its correction, and (5) a gravity meter that will work on board a flying aircraft.

In a preliminary investigation of these problems it appeared that for a 10-mgal accuracy in gravity measurements, existing navigation systems were reasonably adequate. If necessary, accurate ground control techniques could be used, such as optical and radar tracking ranges, Decca, Shoran, and Hiran. Their use would at least permit the surveying of many areas of the world where gravity data are lacking. Experience in flying at high altitudes in large aircraft indicated that under good conditions a smooth flight could be obtained, but the stability of an aircraft and the accelerations encountered in flight were areas which needed further investigation. The Eotvos effect, which is directly con-

nected with the aircraft velocity and hence the navigation system, is of tremendous importance in a fast-flying aircraft and has been a major barrier for gravity measurements in an aircraft. An investigation of this problem, however, led to the conclusion that with proper flight programming, accurate Eotvos corrections could be made, in fact, more precisely than is possible in gravity work at sea. Having thus obtained reasonable surety that four major problems could be handled, there remained the problem of a gravity meter itself. It has long been thought that the development of a new gravimeter employing a new principle would be required for successful measurements in an aircraft. The first logical step, however, was to test an existing submarine or surface-type sea gravity meter on the assumption that if it would work on a ship at sea it could conceivably work on an aircraft under smooth flight conditions. Such a meter was available from LaCoste and Romberg, who had recently developed a model for use on surface boats and had just tested it at sea [LaCoste, 1959]. This same meter was used for the tests in an aircraft.

The tests were done primarily as an investigation or feasibility study and not for actual gravity measurements. However, since the gravity meter actually worked properly and observations were made that constitute the first known measurements, in an aircraft, of the acceleration due to gravity, the results of this program are presented in this report.

Geodetic requirements. The fundamental geodetic problem involving gravity is the determination of the undulations of the geoid and the shape of the earth. For this purpose, average gravity values for $1^\circ \times 1^\circ$ squares (or even $5^\circ \times 5^\circ$ squares) are desired over the entire surface of the earth, but are primarily required over the sea and inaccessible land and ice areas where gravity data are lacking. These gravity values are referred to mean sea level as datum and are required to be accurate to only 10 mgal.

This type of measurement is ideally suited for observations in an aircraft because the value of gravity at a point above the earth is the result of the gravity field over a very large area of the earth's surface beneath the point. This problem has only recently been investigated by W. M. Kaula, C. Tsuboi, E. Tengstrom,

U. Uotila, and others, and papers containing their results will undoubtedly be available in the future. Personal communications from W. M. Kaula and C. Tsuboi give the same results by two different methods. Consider the surface of the earth divided into infinitesimal circular zones of equal width centered about a point on the earth's surface. The contribution of the circular zones increases to a maximum as the zone radius increases from zero. At a radius which is $1/\sqrt{2}$ times the height of the observation point, the maximum contribution for a zone is obtained. For zones of progressively larger radii the contribution decreases until, beyond a zone at a radius of 3 times the observation height, it becomes negligibly small. Therefore, the variations in gravity at a point 30,000 feet above the earth's surface will primarily be the result of the crustal masses over an area on the earth's surface about 40 miles in diameter. In a moving aircraft a gravity 'profile' would be obtained which would be the result of the gravity field along a strip of the earth's surface about 40 miles wide. Since instantaneous readings are not possible with present sea gravity meters, a time average reading must be made. In sea work, for example, 10- to 30-min time averages are taken for one gravity measurement since the distance traveled is not great. In an airborne gravity meter, however, a 5-min (30 mi) or 10-min (60 mi) time average would be desirable which would give an average gravity value over an area such as $1^\circ \times 1^\circ$ square. If an accuracy of 10 mgal could be obtained, the information would be more than adequate.

For this case, the geographical location of the aircraft is not critical because the change in gravity with latitude is only 1 mgal/mi, and a position error of a few miles would not significantly change the average gravity value over a $1^\circ \times 1^\circ$ square. Also, the Eotvos corrections need only be correct to 5 mgal, and the elevation need only be known to about 50 feet. Existing aircraft navigation systems may be expected to give latitude and longitude accurate to $\pm \frac{1}{4}$ miles, track accurate to $\pm \frac{1}{2}^\circ$, and ground speed accurate to ± 1 knot, if the systems work within the manufacturer's specifications and proper flight programming is used. Radar altimeters are available which measure height above terrain or water to 25 feet or

ter. Thus conditions are favorable for meeting this geodetic requirement with an airborne gravity meter.

The Eotvos problem. The term 'Eotvos effect' has arisen from gravity observations made at sea. The east-west component of the ship's velocity adds vectorially to the eastward tangential velocity caused by the earth's rotation, thereby increasing or decreasing the radial acceleration. This change in acceleration in the direction of normal gravity is known as the Eotvos effect.

For an aircraft flying at high velocities the prospects of getting sufficiently accurate Eotvos corrections have never been considered favorable, not only because the corrections are very large, but also because it has been popular opinion that the velocity of the aircraft must be measured to at least 0.1 knots to give an accuracy comparable to that obtained in gravity measurements at sea. An analysis of this problem shows that this opinion is not correct; adequate Eotvos corrections can be obtained by measuring the velocity (ground speed) of the aircraft to only 1 knot, and that under certain flight conditions the high ground speed can actually reduce the error in the Eotvos correction.

For ideal conditions of constant velocity and elevation, the equation for the Eotvos correction is similar to that used for gravity measurements at sea and can be taken as

$$= \frac{R_\phi + h}{R_\phi^2} [(V_\phi + v_e)^2 + v_n^2 - V_\phi^2] \quad (1)$$

where

Δg = Eotvos correction in mgal

R_ϕ = radius of the earth in latitude ϕ

V_ϕ = tangential velocity at the surface of the earth at latitude ϕ

v_e = easterly component of the ground speed of the aircraft

v_n = northerly (or southerly) component of the ground speed of the aircraft

h = elevation above mean sea level

Equation (1) can be simplified to

$$\Delta g = \frac{R_\phi + h}{R_\phi^2} [2V_\phi v_e + v^2] \quad (2)$$

where $v = \sqrt{v_e^2 + v_n^2}$ = ground speed of aircraft.

A term of $(R_\phi + h)/R_\phi$ corrects the velocities from surface values to values at the altitude of the aircraft; the effect is only about 0.1 per cent. Since V_ϕ and R_ϕ can be computed as precisely as necessary, the only sources of error are in v_e , v_n , and v . For an eastbound flight the error in the first term of equation (1) and equation (2) is

$$\frac{\partial(\Delta g)}{\partial v_e} = 2 \frac{R_\phi + h}{R_\phi^2} (V_\phi + v_e)$$

Since h can be considered negligible, this equation reduces to

$$\frac{\partial(\Delta g)}{\partial v_e} = \frac{2}{R_\phi} (V_\phi + v_e) \quad (3)$$

For westbound flight this equation becomes

$$\frac{\partial(\Delta g)}{\partial v_e} = \frac{2}{R_\phi} (V_\phi - v_e) \quad (4)$$

The significance of the error term is illustrated in Figure 1, which shows the change in the radial component V_ϕ^2/R_ϕ of the centrifugal acceleration with the tangential velocity at the surface of the earth. The error equations (3) and (4) are equivalent to the slope of the curve in Figure 1. If $v_e = 0$ knots at the equator, the error or slope is 7.5 mgal/knot. If an aircraft flies eastward so that $v_e = V_\phi/2$ (about 450 knots), the error increases to only 11 mgal/knot, which is not unreasonable. If the aircraft flies westward so that $v_e = -V_\phi/2$, the error decreases to 3.7 mgal/knot. At latitudes north and south of the equator, these error values are even smaller. The v^2 term of equation (2) is represented by the lowest part of the curve in Figure 1, and for velocities up to about 300 knots the error is only about 1 mgal/knot.

This analysis shows, therefore, that, even though the Eotvos corrections are very large (over 1000 mgal), sufficient accuracy can be obtained in the corrections by measuring the ground speed of the aircraft to only 1 knot. The best conditions for making accurate corrections are (1) by flying westward so that the exact track is not critical, in which case the error can be reduced to about 3 mgal/knot, and (2) by flying precisely due north or due south so that only a small correction (about 300 mgal) is necessary. There is, however, no reason why

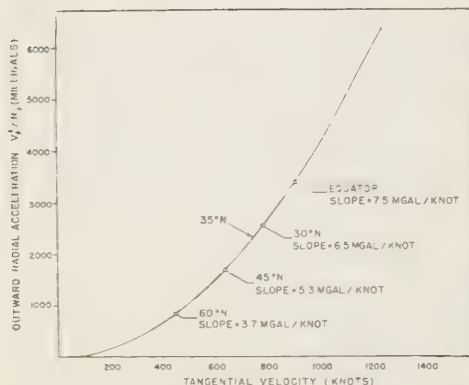


Fig. 1. Curve showing the variation in the radial acceleration V_e^2/R_e with the tangential velocity for points on the surface of the earth. On this scale the effect of the change in the earth's radius is negligible.

flights cannot be made in any direction, since the error in the Eotvos correction would never exceed 7.5 mgal/knot except for flights going due east. Even on an easterly heading, the error per knot does not increase significantly for velocities up to about 200 knots, and an error of 11 mgal/knot for a velocity of about 450 knots is not a serious problem.

The gravity meter. The gravity meter used for these tests was the LaCoste and Romberg Sea Gravity Meter No. 5, which was designed for use on a surface ship. This model is similar to previous submarine meters which have been described by Spiess and Brown [1958] and are currently in use. A description of this meter (No. 5) with its modifications and the results of tests with it at sea on a surface ship has already been presented by LaCoste [1959].

A picture of the gravity meter is shown in Figure 2. The meter itself is suspended from a gimbal joint and therefore always hangs in the resultant vertical when a horizontal acceleration is acting. Two precise accelerometers with 1-min periods measure the horizontal accelerations, and a correction is computed which is accurate to 1 mgal if the periods of the accelerations are considerably shorter than 1 min. The correction is automatically made to the meter reading. The accelerations measured by both accelerometers are continuously recorded.

This surface-type meter has been designed

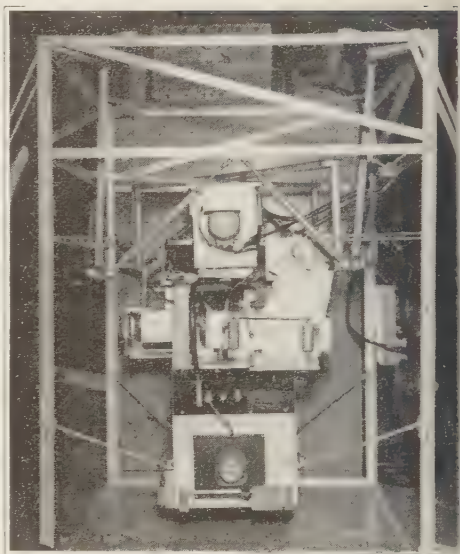


Fig. 2. The LaCoste and Romberg Sea Gravity Meter.

to withstand horizontal and vertical accelerations up to about 100,000 mgal, and it features an accurate averaging method for obtaining readings. Although satisfactory on a ship, this method was too slow to be used on an aircraft and another reading technique was used. As in a normal gravity meter, there is a calibrated dial for changing the spring tension to move the beam to its null position. In order to obtain readings when the beam is in motion and cannot be properly nulled, the differential equation involving the acceleration, velocity, and displacement of the beam must be evaluated. This evaluation is done by an analogue computer and the result, referred to as the 'beam correction,' is applied to the counter reading. The computer in the gravity meter actually gives the integral of the beam correction rather than giving the beam correction itself. The integral is used for convenience in averaging. A pen recorder gives a continuous trace or curve of this integral of the beam correction. By computing corrections from this curve and applying them to the counter readings, which are also recorded continuously, proper average readings can be obtained for any time interval.

The test flights. This experiment was con-

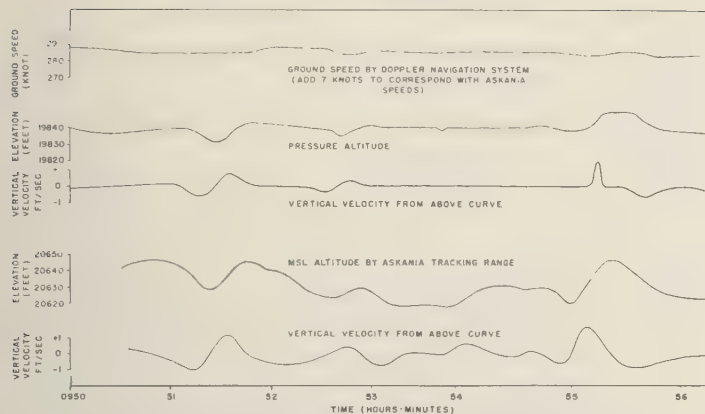


Fig. 3. Flight data from the Askania tracking range and the navigation instruments on the aircraft for northbound flight 1 at 20,000 feet on November 6, 1958.

ected at Edwards Air Force Base, California, with flights on November 6 and 7, 1958. The gravity meter with its associated frame was mounted on the floor of a KC-135 jet tanker at the center of gravity of the aircraft. Flight data were provided by a Doppler navigation system on the aircraft, and precise data were also obtained by making the flights over an Askania camera tracking range at Edwards Air Force Base. Smooth flight conditions were obtained with an automatic pilot. No radar altimeter or airborne profile recorder was used, hence elevation above mean sea level was available from the tracking range data.

Since the gravity meter was designed for surface work, 20,000 feet was the highest elevation at which northbound or southbound flights could be made with the meter on scale. At this altitude flights were made going north, south, and west at about 300 knots. At 30,000 feet only eastbound flights could be made, and observations were obtained for three such flights at ground speeds from about 300 to 370 knots. The flights were made as high as possible to give smooth flight conditions and good aircraft performance.

Flight information. These flights proved that this KC-135 aircraft was a sufficiently stable platform for making gravity measurements with a gravity meter of this type. The automatic pilot maintained a smooth flight, and the accelerations of the aircraft proved to be within

the measuring capabilities of the gravity meter except when the air was turbulent. The horizontal accelerations were small and the corrections were usually less than 30 mgal. However, the periods of the horizontal accelerations were as long as 42 sec. The full magnitude of these long-period accelerations was not measured by the horizontal accelerometers because the accelerometers have a 1-min period and therefore measure accelerations accurately only when the periods of oscillation are considerably shorter than 1 minute. As a result, an additional correction for the reduced accelerometer response had to be computed and applied to the gravity readings. The correction was computed from the known response of the accelerometers and associated electronic circuits to the period of the horizontal accelerations (which were recorded). In all cases but one the correction was 10 mgal or less. On the northbound flight on November 6, 1958, a horizontal acceleration with a 42-sec period was noted and attributed to the hunting of the automatic pilot. Since this approached the 1-min period of the accelerometers, a large correction of -78 mgal was found for this flight. The vertical accelerations of the aircraft were, of course, much more disturbing and caused a continuous and irregular movement of the gravity meter beam. This is to be expected, since a change in elevation, a change in vertical velocity, and a change in forward velocity are recorded by the gravity

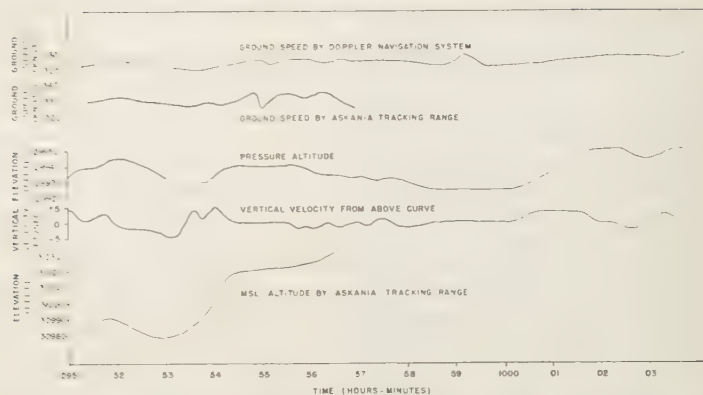


Fig. 4. Flight data from the Askania tracking range and the navigation instruments on the aircraft for westbound flight 2 at 30,000 feet on November 7, 1958.

meter as short-period changes in vertical acceleration or gravity.

Representative curves of the navigational flight data from the instruments on board the aircraft and the Askania tracking range are shown in Figure 3 for northbound flight 1 on November 6, 1958, and in Figure 4 for westbound flight 2 on November 7, 1958. In this work the Askania data were taken to be correct and were used in the computations, and the data from the navigation system were included for comparison. The Askania elevation curves show variations of only about ± 20 feet which represents ± 2 mgal for the change in elevation

only. The pressure altitude measured by an altimeter on the aircraft shows some lag and considerable damping so that the full magnitude of the changes in elevation (as indicated by the Askania elevations) are not recorded. The variations in elevation appear to have periods of about $\frac{1}{2}$ to $1\frac{1}{2}$ min. The vertical velocities created by these changes in elevation are not very large, usually about ± 1 ft/sec and occasionally up to ± 5 ft/sec (Figs. 3 and 4). Since these occur over several tens of seconds, the associated accelerations are not large—usually less than 10,000 mgal—and are easily handled by the meter. On rare occasions (usually at the

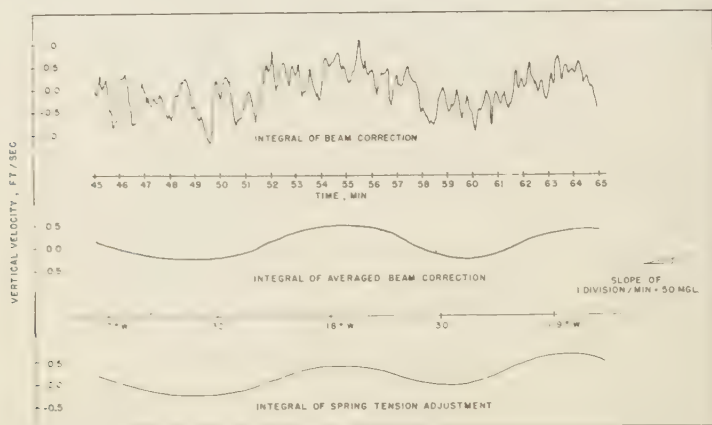


Fig. 5. Gravity data curves for westbound flight 2 on November 7, 1958, illustrating the reduction technique.

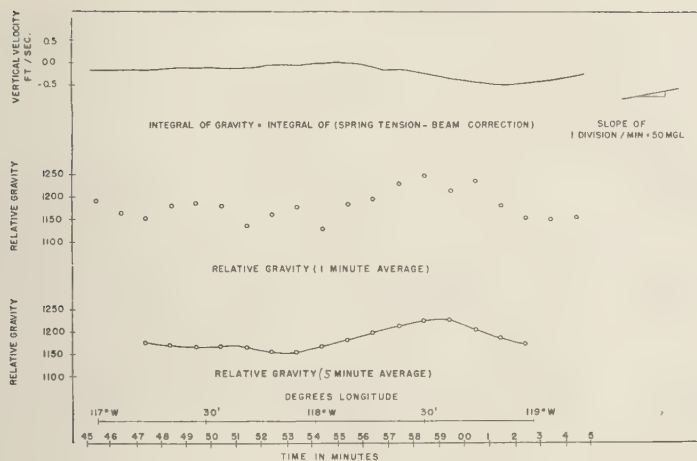


Fig. 6. Gravity data curves for westbound flight 2 on November 7, 1958, illustrating the reduction technique (continued).

part of a run when flight conditions were not settled down) sudden large vertical accelerations occurred which caused the meter beam to hit the stops. The instrument takes about 3 min to recover from such shocks.

Of particular interest is the fact that the ground speed of the aircraft did not remain constant (Figs. 3 and 4) but varied about ± 5 knots with periods from about $\frac{1}{2}$ to 2 min, much the same as the elevation. Although this variation may be considered small, it does, in fact, cause a significant change in the Eotvos effect which affects the gravity meter as a periodic vertical acceleration. For the westbound flights a change in ground speed of 5 knots would cause a change in Eotvos effect of 18 mgal. However, if the ground speed is measured accurately to 1 knot and if an averaging time of several minutes is used, an adequate Eotvos correction can be determined.

During these test flights the gravity meter actually worked satisfactorily. When the meter was in operation, the counter reading was continually changed to keep the beam near the null position. This proved to be unsatisfactory since it complicated the data reduction. It would have been much better to leave the counter reading constant and let the beam position vary with gravity. However, by changing the counter reading, direct in-flight evidence that the in-

strument was affected by gravitational changes was obtained when the counter had to be increased about 100 mgal over the Tehachapi Mountains.

Reduction of relative gravity values. Following a very thorough analysis of the observational data, a reduction technique was evolved which essentially eliminated the effect of all accelerations with periods of about 2 min or less and permitted over-all average corrections for elevation and Eotvos effect to be used. This method is illustrated in Figures 5 and 6, which show the gravity meter data for westbound flight 2 on November 7, 1958. At the top of Figure 5 is the very irregular curve of the integral of the beam correction as recorded by the gravity meter. This curve was averaged by a computer to eliminate the short-period fluctuations and to give the smooth correction curve shown in the middle of the diagram. Much of the long-period variation of this curve was caused by the continual changing of the counter reading mentioned previously. To eliminate this effect, an average counter reading was selected and a curve was computed for the integral of the counter reading (spring tension adjustment). This is shown at the bottom of Figure 5. The difference between the averaged beam-correction curve and the spring-tension curve is therefore caused by changes in gravity. This

TABLE 1. Reduction of Relative Gravity Values for Northbound Flight I, Nov. 6, 1958

Time, PST	A Beam correction curve ordinates, chart divs.	B Spring tension curve ordinates, chart divs.	C Gravity curve ordinates, chart divs.	D Slope, chart divs. per min.	E Counter correction, ($D \times \text{factor}$) dial divs.	F Corrected counter reading, ($E + 226$) dial divs.	G Gravity reading, ($F \times \text{factor}$) mgal	H 5-min. average gravity reading, mgal	I Final corrected gravity reading, mgal	J Average over area $ABCD$, mgal
0946	45.5	49.0	-3.5	5.9	-148.1	72	69			
47	49.1	46.7	2.4	0.2	-5.0	215	207			
48	46.3	43.7	2.6	-1.3	32.6	253	243	171	579	
49	42.2	40.9	1.3	2.1	-52.7	167	161	179	587	
50	42.0	38.6	3.4	1.5	-37.7	182	175	182	590	
51	41.5	36.6	4.9	0.3	-7.5	112	108	179	587	
52	39.9	34.7	5.2	-0.5	12.6	233	224	188	596	
53	37.7	33.0	4.7	-0.7	17.6	238	229	189	597	
54	37.0	33.0	4.0	0.4	-10.0	210	202	214	622	
55	38.0	33.6	4.4	1.3	-32.6	187	180	213	621	
56	42.0	36.3	5.7	-1.0	25.1	245	236	224	632	
57	45.2	40.5	4.7	-0.3	7.5	228	219			
58	49.9	45.5	4.4	-3.0	75.3	295	284			
59	52.1	50.7	1.4							

Average elevation = 20,630 feet.

Average north velocity = 292 knots.

Average east component of velocity = 12 knots.

Average track = 002° true.

Elevation correction to 20,000 feet = 59 mgal.

Correction for north velocity = 355 mgal.

Correction for east velocity = 72 mgal.

Correction for long period horizontal accelerations = -78 mgal.

Total correction = 408 mgal.

TABLE 2. Reduction of Relative Gravity Values for Southbound Flight 2, Nov. 6, 1953

Time, PST	A Beam correction curve ordinates, chart divs.	B Spring tension curve ordinates, chart divs.	C Gravity curve ordinates, chart divs.	D Slope, chart divs. per min.	E Counter correction, (D × factor) dial divs.	F Corrected counter reading, (E + 220) dial divs.	G Gravity reading, (F × factor) mgal	H 5-min. average gravity reading, mgal	I Final corrected gravity reading, mgal	J Average over area ABCD, mgal
1010	53.5	51.0	2.5	1.7	-42.7	177	170			595
11	57.8	53.6	4.2	-1.2	30.1	250	240			
12	58.5	55.5	3.0	0.6	-15.1	205	197	198	601	
13	56.4	52.8	3.6	0.5	-12.6	207	199	206	609	
14	50.3	46.2	4.1	1.1	-27.6	192	185	189	592	
15	43.0	37.8	5.2	0.2	-5.2	215	207	174	577	
16	35.4	30.0	5.4	2.3	-57.7	162	156			
17	32.5	24.8	7.7	3.7	-92.9	127	122			
18	40.4	29.0	11.4							
Average elevation = 20,675 feet. Average south velocity = 330 knots. Average west component of velocity = 17 knots. Average track = 183° true.										Elevation correction to 20,000 feet = 63 mgal. Correction for south velocity = 454 mgal. Correction for west velocity = -106 mgal. Correction for long period horizontal accelerations = -8 mgal. Total correction = 403 mgal.

TABLE 3. Reduction of Relative Gravity Values for Westbound Flight 3, Nov. 6, 1958

Time, PST	A Beam correction curve ordinates, chart divs.	B Spring tension curve ordinates, chart divs.	C Gravity curve ordinates, chart divs.	D Slope, chart divs. per min.	E Counter correction, ($D \times \text{factor}$) dial divs.	F Corrected counter reading, ($E + 2000$) dial divs.	G Gravity reading, ($F \times \text{factor}$) mgal	H 5-min. average gravity reading, mgal	I Final corrected gravity reading, mgal	J Average over area <i>EFGH</i> , mgal
1105	43.1	48.8	-5.7	-2.2	55.2	2055	1977			
06	41.0	48.9	-7.9	-3.5	87.9	2088	2009	2018	599	599
07	40.1	51.5	-11.4	-4.7	118.0	2118	2038			
08	39.5	55.6	-16.1	-5.1	128.0	2128	2047			
09	39.0	60.2	-21.2							

Average elevation = 20,615 feet.
 Average west velocity = 300 knots.
 Average track = 271° true.

Elevation correction to 20,000 feet = 58 mgal.
 Correction for west velocity = -1475 mgal.
 Correction for long period horizontal accelerations = -2 mgal.
 Total correction = -1419 mgal.

TABLE 4. Reduction of Relative Gravity Values for Westbound Flight 1, Nov. 7, 1958

Time, PST	A Beam correction curve ordinates, chart divs.	B Spring tension curve ordinates, chart divs.	C Gravity curve ordinates, chart divs.	D Slope, chart divs. per min.	E Counter correction, ($D \times$ factor) dial divs.	F Corrected counter reading, ($E + 1400$) dial divs.	G Gravity reading, ($F \times$ factor) mgal	H 5-min. average gravity reading, mgal	I Final corrected gravity reading, mgal	J Average over area <i>EFGH</i> , mgal
0913	40.6	89.9	-49.3	4.1	-103.0	1297	1247			-327
14	42.3	87.5	-45.2	7.3	-183.0	1217	1170			
15	47.4	85.3	-37.9	5.8	-146.0	1254	1206	1199	-309	
16	50.9	83.0	-32.1	5.8	-146.0	1254	1206	1181	-327	
17	53.8	80.1	-26.3	7.5	-188.0	1212	1165	1177	-331	
18	58.0	76.8	-18.8	7.8	-196.0	1204	1158	1160	-348	
19	61.6	72.6	-11.0	8.2	-206.0	1194	1148	1189	-319	
20	62.3	65.1	-2.8	5.0	-126.0	1274	1225			
21	58.2	56.0	2.2	4.0	-100.0	1300	1250			
22	55.0	48.8	6.2							

Average elevation = 31,000 feet.
 Average west velocity = 333 knots.
 Average track = 272° true.

Elevation correction to 30,000 feet = 94 mgal.
 Correction for west velocity = -1592 mgal.
 Correction for long period horizontal accelerations = -10 mgal.
 Total correction = -1508 mgal.

TABLE 5. Reduction of Relative Gravity Values for Westbound Flight 2, Nov. 7, 1958

Time, PST	A Beam correction curve ordinates, chart divs.	B Spring tension curve ordinates, chart divs.	C Gravity curve ordinates, chart divs.	D Slope, chart divs. per min.	E Counter correction, ($D \times$ factor) dial divs.	F Corrected counter reading, ($E + 1220$) dial divs.	G Gravity reading, ($F \times$ factor) mgal	H 5-min. average gravity reading, mgal	I Final corrected gravity reading, mgal	J Average over area $EFGH$, mgal
0945	45.8	49.7	-3.9	-0.6	15	1235	1188			
46	43.2	47.7	-4.5	0.5	-13	1207	1161			
47	41.2	45.2	-4.0	1.0	-25	1195	1149	1171	-327	
48	40.0	43.0	-3.0	-0.1	3	1223	1176	1169	-329	
49	39.0	42.1	-3.1	-0.3	8	1228	1181	1163	-335	
50	38.8	42.2	-3.4	-0.2	5	1225	1178	1165	-333	
51	40.9	44.5	-3.6	1.7	-43	1177	1132	1165	-333	
52	46.1	48.0	-1.9	0.6	-15	1205	1159	1154	-344	
53	50.5	51.8	-1.3	-0.1	3	1223	1176	1155	-343	
54	52.3	53.7	-1.4	1.9	-48	1172	1127	1168	-338	
55	54.5	54.0	0.5	0.3	8	1228	1181	1181	-317	
56	53.3	53.1	0.2	-0.9	23	1243	1195	1196	-302	
57	49.3	50.0	-0.7	-2.3	58	1278	1229	1213	-285	
58	44.0	47.0	-3.0	-3.1	78	1298	1248	1224	-274	
59	40.0	46.1	-6.1	-1.7	43	1263	1214	1221	-277	
1000	39.0	46.8	-7.8	-2.5	63	1283	1234	1206	-292	
01	40.6	50.9	-10.3	-0.3	8	1228	1181	1186	-312	
02	45.3	55.9	-10.6	0.9	-23	1197	1151	1174	-324	
03	49.3	59.0	-9.7	1.0	25	1195	1149			
04	50.8	59.5	-8.7	0.7	-18	1202	1156			
05	50.8	58.8	-8.0							

-303

Average elevation = 30,990 feet.
Average west velocity = 331 knots.
Average track = 271° true.

Elevation correction to 30,000 feet = 93 mgal.
Correction for west velocity = -1585 mgal.
Correction for long period horizontal accelerations =

TABLE 6. Reduction of Relative Gravity Values for Westbound Flight 3, Nov. 7, 1958

Time, PST	A Beam correction curve ordinates, chart divs.	B Spring tension curve ordinates, chart divs.	C Gravity curve ordinates, chart divs.	D Slope, chart divs. per min.	E Counter correction, ($D \times$ factor) dial divs.	F Corrected counter reading, ($E + 1360$) dial divs.	G Gravity reading, ($F \times$ factor) mgal	H 5-min. average gravity reading, mgal	I Final corrected gravity reading, mgal	J Average over area $EFGH$, mgal
1041	36.9	70.3	-33.4	1.9	-48	1312	1262			-330
42	41.1	72.6	-31.5	2.5	-63	1297	1247			
43	46.0	75.0	-29.0	-0.5	13	1373	1320	1281	-342	
44	48.0	77.5	-29.5	-0.9	23	1383	1330	1290	-333	
45	49.5	79.9	-30.4	2.5	-63	1297	1247	1299	-324	
46	54.3	82.2	-27.9	0	0	1360	1308	1304	-319	
47	57.9	85.8	-27.9	0.7	-18	1342	1290			
48	62.8	90.0	-27.2	-1.5	38	1398	1344			
49	64.0	92.7	-28.7							

Average elevation = 30,930 feet.
 Average west velocity = 368 knots.
 Average track = 272° true.

Elevation correction = 87 mgal.
 Correction for west velocity = -1705 mgal.
 Correction for long period horizontal accelerations = -5 mgal.
 Total correction = -1623 mgal.

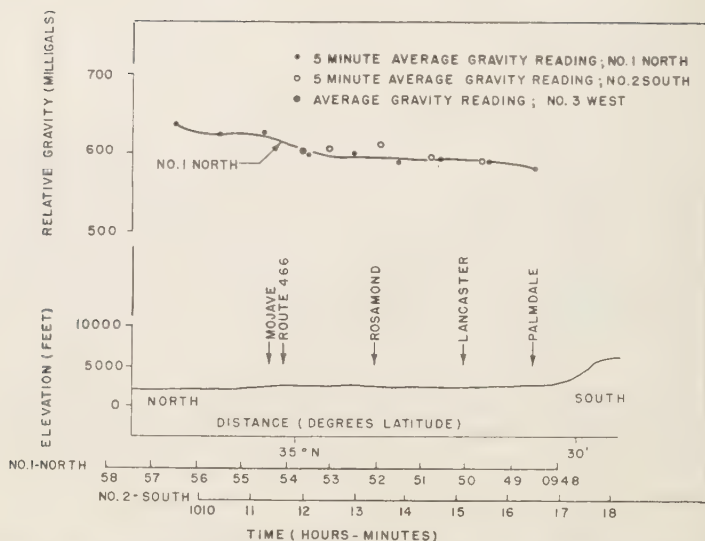


Fig. 7. Relative gravity profile for the northbound and southbound flights at 20,000 feet in relation to a topographic section.

difference curve is the integral of the gravity reading and is shown at the top of Figure 6. Taking slopes along this curve for 1-min intervals gives gravity readings which are plotted in the middle curve. These readings show considerable scattering, which is to be expected from the observed variations in vertical acceleration. By averaging the slope of the upper curve over 5-min intervals more precise gravity values are obtained, which form the lower curve in Figure 6. Since this curve is considerably smoother than the 1-min curve, the conclusion is that a 5-min averaging time is reasonable for air work.

The accuracy of the gravity readings depends on the accuracy of the slope measurements since the counter reading is precisely calibrated. For a 1-min interval the slope can be measured from the record to an accuracy of about ± 5 mgal. For a 5-min interval the slope can be determined to an accuracy of about ± 0.5 mgal. In the reduction of the results, determinations were made for 1-min intervals, and these were then averaged for 5-min intervals. Although this is equivalent to taking slopes for a 5-min interval, the error in the results is likely to be more than ± 0.5 mgal but should not be greater than 1 or 2 mgal. A complete numerical reduction

for all flights is given in Tables 1 through 6. The steps leading to 5-min average gravity readings are identified in columns A to H.

Since the reduction technique essentially eliminates the effect of short-term variations in elevation, ground speed, track, and vertical accelerations, the necessary corrections to the gravity readings are computed using only the average elevation, velocity, and track for each flight. The flight data and computed corrections are tabulated at the bottoms of Tables 1 through 6. The correction for elevation corrects the gravity readings to 20,000 feet for the flights made on November 6 and to 30,000 feet for the flights made on November 7. A free-air correction of 0.09406 mgal/ft was used. The Eotvos corrections were computed with the equation quoted previously. In the northbound and southbound flights the track was not exactly along the meridian and an east-west component of velocity (about 15 knots) was present. As in gravity work at sea, this relatively small component introduced corrections of about 100 mgal. The error in these corrections may be large, since the flight track was taken to the nearest whole degree only. If the track is in error by $\frac{1}{2}^\circ$, which could be the case, an 18 mgal error would result. The sum of all the

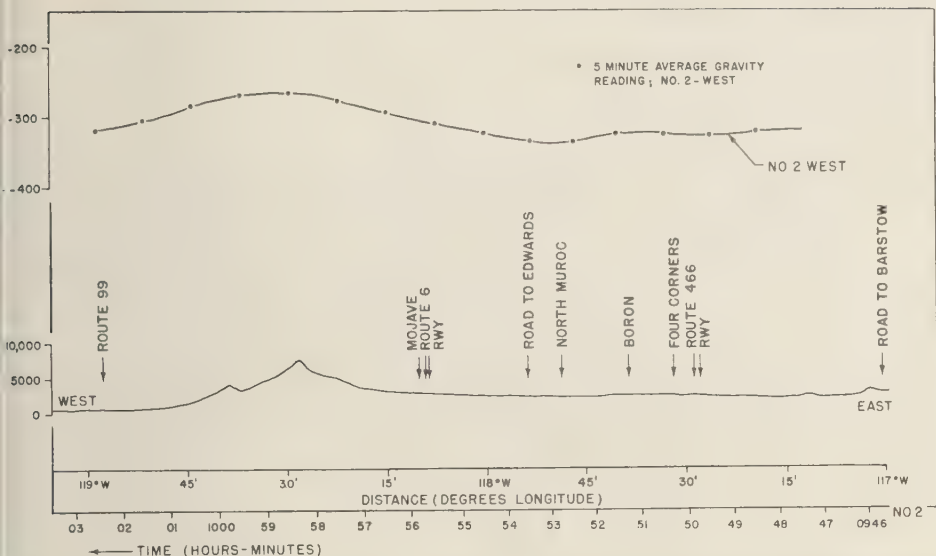


Fig. 8. Relative gravity profile for westbound flight 2 at 30,000 feet in relation to a topographic section through the Tehachapi Mountains.

average corrections for each flight has been added to the 5-min readings in column *H* to give the final corrected gravity readings listed in column *I*. These, then, are the relative gravity values for 20,000 or 30,000 feet, as the case may be.

Relative gravity profiles. The gravity values in column *I* of Tables 1, 2, and 3 for the northbound, southbound and westbound flights at 30,000 feet are plotted in Figure 7 over a north-south topographic section. A smooth curve is drawn primarily through the points for the northbound flight. Although the points show some scatter, especially for the southbound flight, they are for the most part within a ± 2 -mgal limit. The surface gravity in the area is relatively constant except for the change with altitude. The slight increase in gravity at the northern end of the profile may be associated with the Tehachapi Mountains, which are close to the flight line north of the town of Mojave. Whereas the over-all slope of the curve can conceivably be taken at the extreme limits of .9 mgal/mi and 1.2 mgal/mi, the best fit gives a slope of 1.1 mgal/mi. This is the same as the computed latitude effect and is therefore a significant result.

Figure 8 shows the relative gravity profile obtained on westbound flight 2 at 30,000 feet together with a topographic cross section through the Tehachapi Mountains. If the smoothness of the profile is used as a criterion, the observations appear to be accurate to better than 10 mgal. This indicates a good quality of measurement by the gravity meter and a good reduction technique, since the corrections for elevation and Eotvos effect are constant factors and do not affect the relative change from point to point. Of particular interest is the increase in gravity of about 50 mgal directly over the Tehachapi Mountains. Since the plotted values are 5-min averages, the peak value is considerably reduced by the effect of the lower values on the flanks of the anomaly. The 1-min average values in Table 5 give a peak anomaly of about 120 mgal, but these values are not considered accurate enough for correlation with the mountainous topography.

Of the three westbound flights, only flight 2 gave good gravity results for any length of time. On the other flights observations were acceptable only for short intervals. The results of flight 1 are scattered because the meter was recovering from a severe acceleration which caused the

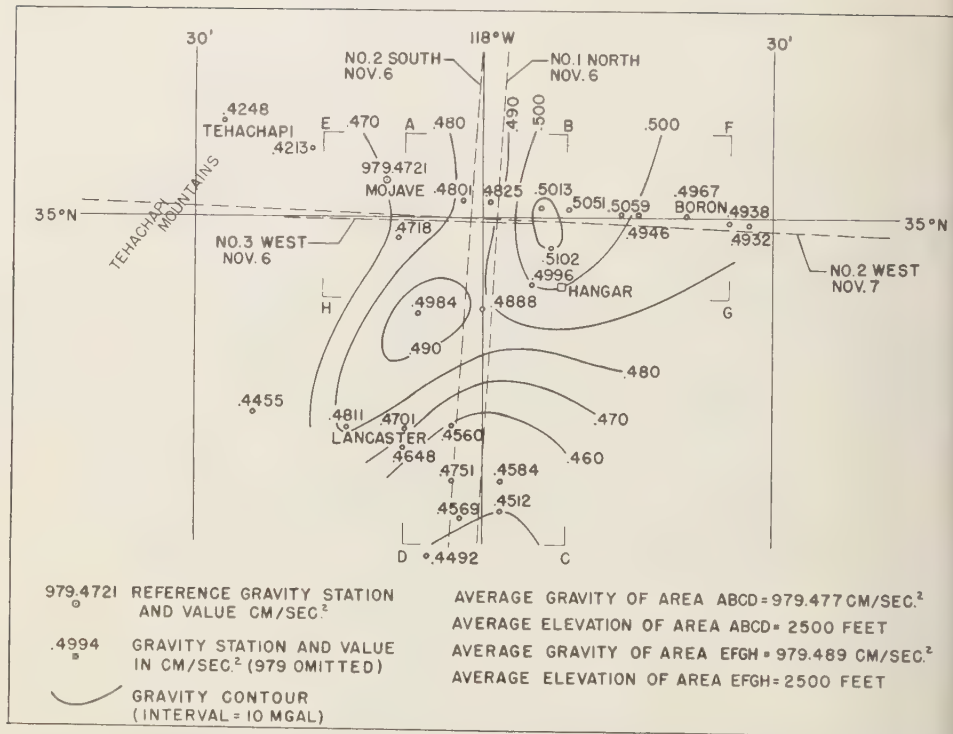


Fig. 9. Surface gravity in the vicinity of Edwards AFB, California. Values are observed gravity with no adjustment for differences in station elevation. The gravity contours, therefore, reflect primarily the surface topography.

TABLE 7. Comparison of Gravity Values at Elevation Computed from Aerial Measurements and Ground Measurements

Item	20,000 ft, Area ABCD	30,000 ft, Area EFGH
<i>Aerial measurements</i>		
Average gravity reading at elevation	595 mgal	-330 mgal
Gravity reading at hangar	2,276 mgal	2,276 mgal
Gravity difference	-1,681 mgal	-2,606 mgal
Gravity value at hangar	979,507 mgal	979,507 mgal
Average gravity value at elevation	977,826 mgal	976,901 mgal
<i>Ground measurements</i>		
Average gravity value at surface	979,477 mgal	979,489 mgal
Average elevation	2500 ± 100 ft	2500 ± 100 ft
Free-air correction	-1,646 mgal	-2,587 mgal
Average gravity value at elevation	977,831 mgal	976,902 mgal

am to hit the stops. The results of flight 3 are not too reliable for the same reason; also this portion of the flight was beyond the tracking range and precise control data is lacking.

Gravity values at elevation. The previous profiles are simply relative gravity in which the change from point to point is meaningful, but there is no gravity datum. Since the gravity meter was read on the ground at the hangar prior to the test flights, gravity values can be determined for the gravity readings in column *J* of Tables 1 through 6. The basic gravity data on the ground are, however, rather inadequate, and only an approximate comparison has been attempted.

The surface gravity in the vicinity of Edwards Air Force Base is shown in Figure 9. Most of the stations were established by a land meter at the time of the flight tests and no elevations are available. A gravity station previously established by G. P. Woollard (personal communication) at the town of Mojave was used as a reference. The values plotted are the observed gravity values at the stations, and therefore the contours reflect primarily the changes in elevation in the area together with the latitude effect. The tracks of various flights are shown, as well as latitude and longitude lines, so that the relative positions of the gravity profiles and topographic sections in Figures 7 and 8 can be compared.

The ground data do not cover sufficient area to permit a complete study of the aerial measurements; therefore the correlation has been limited to the areas *ABCD* and *EFGH* shown in Figure 9. Average gravity readings over these areas for the flights at 20,000 feet are given in column *J* of Tables 1 through 3. Similarly, average readings over area *EFGH* for the flights at 20,000 feet are given in Tables 4 through 6. In Table 7 these average readings are evaluated relative to a known gravity value at the hangar on the ground. The average gravity at 20,000 feet over area *ABCD* and at 30,000 feet over area *EFGH*, are also computed using the average surface gravity and a free-air correction (0.09406 mgal/ft). A comparison of the results shows excellent agreement between the gravity values obtained by the two different methods for each elevation. This is rather fortuitous in view of the possible errors in the gravity meas-

urements themselves, the average surface gravity values (including the fact that the width of the areas is too small), the average elevations of the two areas, etc. However, this comparison does indicate that the measured average gravity values are correct to about 10 mgal and that the average gravity measured in the air over an area can be correlated with the average surface gravity of that area.

General conclusions. The tests have proved that aerial gravity measurements are feasible and can be performed with an existing type of sea gravity meter with an accuracy comparable to that for gravity work done at sea. It must be remembered that this experiment was done as a feasibility study and there were limitations in the meter, flight conditions, and observational procedure. The results have, therefore, been analyzed primarily with consideration for effects of the order of 10 mgal and not necessarily for effects of 1 or 2 mgal.

Although not of a precise nature, the actual measurements of the acceleration due to gravity are the first known to have been made in a flying aircraft and appear suitable for the intended geodetic application. The smooth relative gravity profiles, the measured change in gravity with latitude, and the detected gravity anomaly over the mountain range indicate an accuracy of measurement of the order of 10 mgal. The average gravity values over the test area, measured at 20,000 and 30,000 feet, are consistent with the average gravity over the same area on the ground if a simple free-air correction is made. This supports the possibility of determining an average gravity value in the air over an area such as $1^{\circ} \times 1^{\circ}$ square and reducing this to a sea-level value which is, indeed, the geodetic requirement.

The KC-135 aircraft, flying at high altitudes under control of an automatic pilot, proved to be a sufficiently stable platform from which to make gravity measurements with a sea-type gravity meter. Although the reference flight data for this experiment were obtained by the Askania tracking system, there is evidence that satisfactory flights can be made with a Doppler navigation system and a radar altimeter on the aircraft. The conclusions of the investigation of the Eotvos problem were well verified by the results of these tests, in that there was excel-

lent agreement in the reduced data for flights which were made on different headings and with different ground speeds. The technique that was developed for the reduction of the results is important because it shows that the effect of periodic changes in elevation, vertical velocity, and ground speed can be eliminated from the gravity measurements. This means that relative gravity measurements can be made very precisely for observations along a profile. For true gravity values, the proper elevation and Eotvos corrections can be determined by using only the average elevation and the ground speed. These can be determined accurately for a flight if continuous readings are taken.

Whereas the sea gravity meter operated satisfactorily on these tests, it did not have the characteristics of range, damping coefficient, accelerometer period, measuring system, etc., which are needed for more advanced tests. However, this work has indicated many desirable modifications and new features which can be added to the existing model for use in an aircraft, and this would be the first airborne gravity meter. When such a meter is available, more precise and reliable aerial gravity work will be forthcoming, and perhaps the actual application of gravity measurements in an aircraft will be possible in the not too distant future.

Acknowledgments. The opportunity to perform this work was made possible by Owen Williams, Chief, Geodesy and Gravity Branch, Geophysics Research Directorate, who had the foresight to initiate the airborne gravity meter program as a project task and who gave wholehearted and untiring support to the work. The success of these tests was in part due to the excellent cooperation and support of the various participating groups. The authors extend their appreciation to Lt. Bruce C. Murray of the Geophysics Research Directorate who assisted in arranging the tests and participated in the flight; the Space Positioning Branch, Askania and Radar Sections, Edwards AFB; and the Data Reduction Center, Edwards AFB.

REFERENCES

- Fedynskiy, V. V., Development of apparatus for conducting gravimetric measurements while in motion, *Izvest. Akad. Nauk SSSR, Ser. Geofiz.* no. 1, 146-152, 1959.
- LaCoste, L. J. B., Surface ship gravity measurements on the Texas A and M College ship the "Hidalgo," *Geophysics*, 24, 309-322, 1959.
- Lundberg, H., Airborne gravity surveys, *Bull. Can. Inst. Met.*, p. 465, Aug. 1957.
- Spiess, F. N. and G. L. Brown, Tests of a new submarine gravity meter, *Trans. Am. Geophysical Union*, 39, 391-395, 1958.

(Manuscript received August 7, 1959.)

The Use of Love Waves for the Study of Earthquake Mechanism¹

KEIITI AKI²

*Seismological Laboratory
California Institute of Technology
Pasadena, California*

Abstract. Long-period Love waves of continental path were successfully used for the study of the mechanism of earthquakes. Wave forms were compared for earthquakes of similar size which occurred within a limited area and were recorded by the same instrument at a given station. The aftershocks of the Kern County earthquake of 1952 and the series of Nevada shocks during 1954 were studied, and the direction of lateral fault motion derived from Love waves was found to agree perfectly with that derived from the data on body waves. We also found a very definite fact which cannot be explained by postulating a single couple for an earthquake source and which strongly supports Honda's hypothesis of a pair of couples. This additional information from the Love wave data confirmed the belief that the fault plane solutions have great geotectonic significance, at least for the area studied.

Introduction. The mechanism of earthquakes has been studied extensively [Hodgson, 1957]. In these studies, the sense of the initial pulse of P waves is used for the determination of the nodal planes, and the amplitude and the sense of S waves are sometimes used for the determination of the fault plane. So far, the sense of the surface waves has never been studied for this purpose except in *Tukey's* stimulating proposal [1959]. Since body waves, especially at their very beginning, usually are short-period waves, it is not impossible that the mechanism deduced from them represents only the minor motion at the very beginning of the faulting and does not present its general feature. Therefore, any additional information about the source of surface waves of long period will be important for a better understanding of the phenomena at the source.

Besides, since the information from S waves is usually not reliable, the sense of Love wave motion (which must have a similar radiation pattern to that of S waves, though not identical) will be very useful in determining which of the two nodal planes is the actual fault plane. The uncertainty in the fundamental assumption on the mechanism (that is, a single couple or a

pair of couples) may also be resolved from this source of information.

Data. In this preliminary report, we have dealt with the Love waves from the aftershocks of the Kern County, California, earthquake of 1952 and those from the series of shocks which occurred in Nevada during 1954. The fault plane solutions for most of these shocks have already been given by various authors from the P wave data. The main shock in Kern County was studied by *Gutenberg* [1955], and the aftershocks by *Båth and Richter* [1958]. Some of the Nevada shocks were studied by *Tocher* [1955] and *Romney* [1957].

The earthquakes studied here are listed in Table 1, with the type of fault motion and the direction of the strike. The number attached to the Kern County shock is the serial shock number assigned by *Richter* [1955].

The records of the Kern County aftershocks were gathered at the Seismological Laboratory in Pasadena, California, from almost all the stations of the world by Dr. Beno Gutenberg. It was found that Love waves of well-dispersed form, varying in period from about 30 sec to 8 sec, were recorded at the stations located in the northern and northeastern parts of North America. The records of Ottawa, Weston, Palisades, and Resolute Bay are used in this study. Unfortunately, stations in other parts of North America did not record Love waves of long period, and stations outside North America were

¹ Contribution No. 937, Division of Geological Sciences, California Institute of Technology.

² Visiting Research Fellow on leave from Earthquake Research Institute, Tokyo, Japan.

TABLE 1. List of Earthquakes

No.	Date	Origin Time, GCT		Epicenter				<i>M</i>	Direction of Strike, deg	Type of Fault	Reference
				W		N					
		hr	min	deg	min	deg	min				
1952											
75	July 23	00	39	118	35	35	22	6.1	N 50 E	Dip slip	<i>Báth & Richter</i>
89	July 23	13	17	118	49	35	13	5.7	(N 80 E)	Dip slip	[1958] “
117	July 25	19	10	118	30	35	19	5.7	N 45 E	Left-hand strike slip	“
118	July 25	19	43	118	30	35	19	5.7	N 31 E	Left-hand strike slip	“
141	July 29	07	04	118	51	35	23	6.1	N 53 E	Right-hand strike slip	“
194	Aug. 22	22	41	118	55	35	20	5.8	N 49 E	Right-hand strike slip	“
1954											
	July 6A	11	13	118	32	39	25	6.8	N 24 W	Right-hand strike slip	<i>Tocher</i> [1955]
	July 6B	22	08	118	30	39	18	6.0	—
	Aug. 31	22	21	118	30	39	30	6.5
	Dec. 16	11	07	118	07	39	17	7.1	N 11 W	Right-hand strike slip	<i>Romney</i> [1957]

too distant to record clearly the waves of these aftershocks. For the Nevada shocks, the records of Palisades and Pasadena were available. The instrument constants applicable to this study are listed in Table 2.

Method. The observed wave forms of the Love waves will depend on (1) the mechanism of the earthquake, (2) the azimuthal direction toward the station from the source, (3) the magnitude of the earthquake, (4) the epicentral distance, (5) the structure of the earth's crust along the wave path, and (6) the response characteristic of the recording instrument.

In this paper we are concerned only with the comparison of earthquakes of similar size in a limited area as recorded by a single instrument

at a given station. Therefore, (2), (3), (5), and (6) of the above factors are common to all the seismograms to be compared. The effect of factor (4), the epicentral distance, was found to be negligible in the cases we dealt with. If the epicentral distance differs by Δx between two earthquakes, the amount of the difference in wave form measured by the difference in the length of time from the maximum period crest to the minimum period crest can roughly be estimated to be $\Delta x(1/C_{\min} - 1/C_{\max})$, where v is the phase velocity.

In the case of the Kern County aftershocks, for example, the maximum difference in the epicentral distance was 20 miles. The calculation based on a reasonable model of crustal structure

TABLE 2. Instrument Constants

Station	Type of Instrument	<i>T_s</i> , sec	<i>T_g</i> , sec	Damping
Weston	Benioff long-period (N-S)	1	75	
Ottawa	Milne-Shaw (N-S)	12.0		20:1
Resolute Bay	Sprengnether (E-W)	16.0	16.0	Critical
Pasadena	Strong Motion (E-W)	8.0		16:1
Palisades	Benioff long-period (N-S)	1	75	$\epsilon_s = 1.5\omega_s$ $\epsilon_g = 6\omega_g$

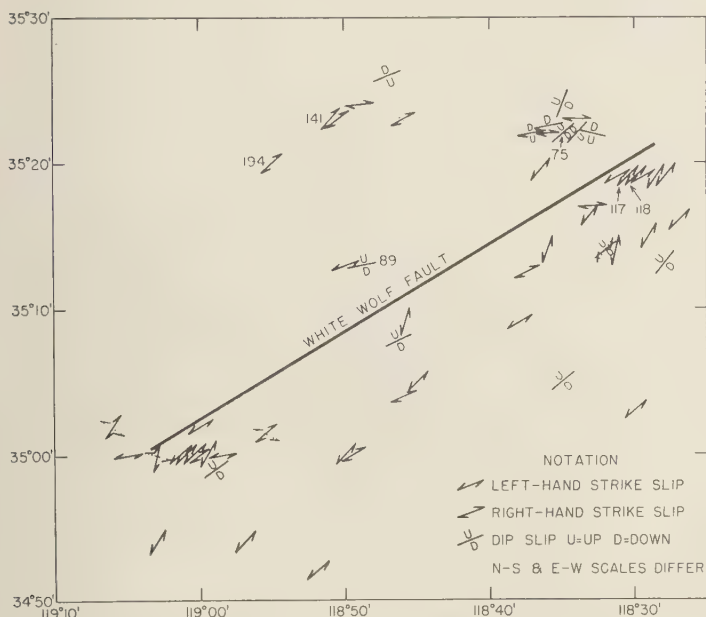


Fig. 1. Map of the Kern County aftershock area showing the individual solutions. (After M. Båth and C. F. Richter).

shows that the corresponding difference in the length of time from the crest of the 30-sec period to the crest of the 8-sec period is about 1 sec. This amount is negligible in our study.

Since we are concerned with the long-period waves, the motion at the source can safely be assumed to be an impulse. Thus, for similar types of fault motion, the only difference expected in the wave form is the difference in clarity. In the case of the Ottawa, Weston, and Resolute Bay seismograms, comparison of the original seismograms by eye showed very clearly whether the sense of their motions was the same or opposite. On the Palisades records it was necessary to remove the second-mode waves by digital filtering before comparing the seismograms.

Kern County aftershocks. The results of the surface wave method were compared with the fault plane solutions obtained by Båth and Richter [1958] from the P wave data. Figure 1, reproduced from their paper, shows on a map the direction of strike and the type of fault motion for each of the aftershocks. The earth-

quakes studied in the present paper are indicated by arrows with the shock numbers assigned by Richter.

The fault plane given by Gutenberg [1955] for the main shock has the direction of strike N 50° E, dipping 60° to 66° toward E 50° S (Fig. 2). The slip direction was upward and northward in the southeastern block relative to the northwestern block. The dip-slip component was about 1.4 times the strike-slip component. Unlike the main shock, the aftershocks were found by Båth and Richter to exhibit considerable strike slip. Left-hand strike slip dominates both along the White Wolf fault and to the south of it. Right-hand strike slip and dip slip occur to the north of it as shown in Figure 1.

The records of the Love waves for earthquake 194 are shown in Figure 3a. The stations are Ottawa, Weston, and Resolute Bay, their locations being shown in Figure 2. These waves are identified as Love waves by their polarization and arrival times. The records are used as standard wave forms with which the wave forms of the other earthquakes are correlated. Accord-

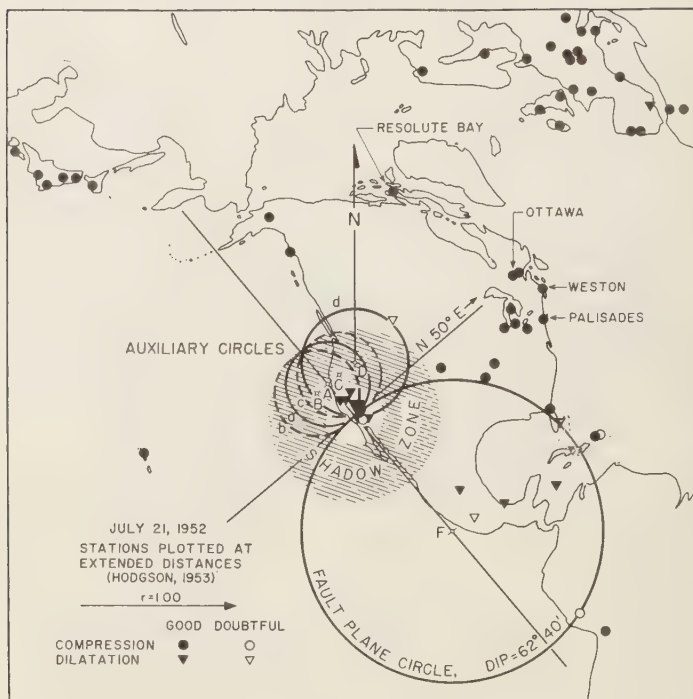


Fig. 2. Observed compressions and dilatations in the main earthquake of July 21, 1952. (After B. Gutenberg).

ing to Båth and Richter, this earthquake is right-hand strike slip, as is shown in Figure 1 and Table 1.

Earthquake 141 is also right-hand strike slip. Figure 3*b* shows the Love waves from this earthquake. Since earthquake 141 is bigger than earthquake 194, the amplitude is larger and the longer-period waves appear in the early portion. But, if one arranges these two seismograms as shown in the figure, one finds that the coincidence in the times of the crests and troughs between the two earthquakes is excellent for each station. If the sense of one of the seismograms is reversed, it is not possible to obtain good coincidence between the seismograms.

Earthquakes 117 and 118 are reported to have occurred at the same place, earthquake 118 following 117 by about 30 minutes. These two seismograms are almost identical in detail. Both are found by Båth and Richter to be left-hand strike slip with a slight difference in the direction of the strike. Figure 3*c* shows the Love

waves from earthquake 117, for which one cannot find any good crest-to-crest or trough-to-trough correspondence with earthquake 194. The correspondence, however, between the crests for this earthquake and the troughs for earthquake 194 is excellent, as is shown in the figure. Therefore, their senses are opposite, as is expected from the P wave data.

Earthquakes 75 and 89 are reported as being dip slip. These earthquakes, however, also generated Love waves, as is shown in Figures 3*d* and 3*e*. The wave form for earthquake 75 fits that for earthquake 194 very well; therefore, the motion at the source had a right-hand strike-slip component. This is consistent with the fact that all the shocks of strike-slip type which occurred close to the source of this earthquake are right hand (Fig. 1). On the other hand, the wave form for earthquake 89 correlates closely with that for earthquake 117. The motion at the source, therefore, had a left-hand strike-slip component. This is consistent with the

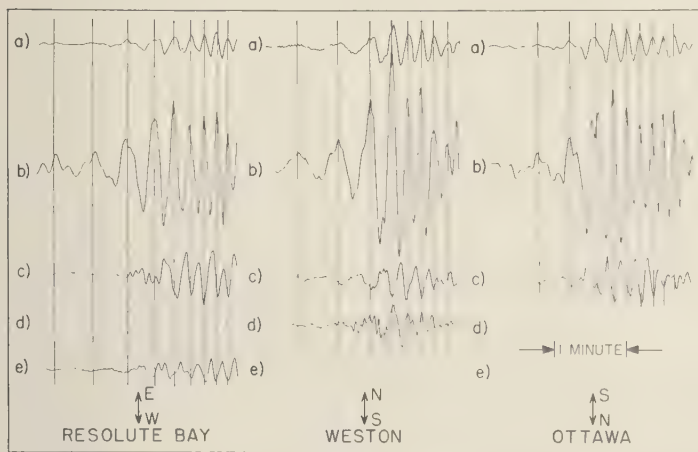


Fig. 3. Love waves from the Kern County aftershocks recorded at Ottawa, Resolute Bay, and Weston.

motion of a nearby earthquake for which Båth and Richter determined left-hand strike-slip motion (Fig. 1). The blanks in Figure 3 simply mean that we do not have those records.

Since Båth and Richter did not use S wave data, they had to choose, on some basis, one of two alternative nodal lines as the actual fault line in the case of strike slip. They chose the one which is more nearly parallel to the White Wolf fault. Therefore, the agreement we found between the mechanisms from the P wave and the Love wave data confirms their interpretation that the mechanisms of these aftershocks were not random in space, but that they imply a definite geotectonic significance. The conclusion of Båth and Richter concerning the difference in mechanism between the main energy release and the secondary release is strongly supported.

Nevada shocks. The eastern part of Churchill County, Nevada, was quite active seismically from July to December, 1954. The largest earthquake, ($M = 7.1$) occurred just north of Fairview Peak on December 16, 1954. Romney [1957] studied the mechanism of this earthquake from the P wave data and showed that the seismic solution agrees well with the geodetic measurement which was made just before and after the occurrence of this earthquake. His solution gives a fault striking $N 11^\circ W$ and dipping $62^\circ E$. The strike-slip component, which is right hand, is about twice the dip-slip component.

Tocher [1955] gave the fault plane solution for

two of those Nevada shocks, one of July 6A, 1954, and the other of August 24, 1954. Two alternative nodal planes for each of these shocks are given in Hodgson's [1957] catalog. If we take for each of the shocks the nodal line which is more nearly parallel to that of the Fairview Peak earthquake, the fault motions for both earthquakes must be right-hand strike slip.

The records of the Love waves by a strong-motion seismograph at Pasadena were available for the Fairview Peak earthquake of December 16, 1954, and the earthquake of July 6A. Since the Love waves are not well enough developed to permit a wave form correlation, because of the short epicentral distance (about 590 km), two records are arranged in Figure 4 so that the arrival times of the initial P waves found from the records of a short-period seismograph coincide with each other. This comparison is allowable because the difference in the epicentral distance between the two is small (about 10 km). The coincidence between the two wave forms in the long-period range assures us that the fault motions of both earthquakes have the same direction in strike-slip component, that is, right hand.

Palisades records of the Love waves by the long-period Benioff seismograph were available for the following earthquakes (Table 1):

1. 1954 July 6A	1113 GCT
2. 1954 July 6B	2207 GCT
3. 1954 August 31	2220 GCT

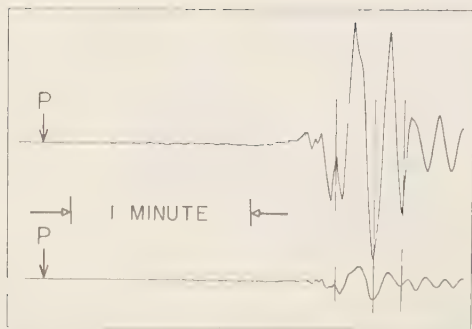


Fig. 4. Love waves from the Nevada shocks of December 16 and July 64 recorded at Pasadena.

The record of earthquake July 64 is the one shown by *Ewing and others* [1957] as a good example of continental Love waves. A wave of period of more than 100 sec can be recognized in this record. However, the long-period waves for the other two earthquakes were masked by shorter-period waves which may be Love waves of the second mode. According to *Dorman's* [1959] calculation on Love waves for various models of the layered crust, the predominant period of the second mode at the early portion of the Love waves is nearly equal to H/β_1 , where H is the thickness of the earth's crust and β_1 is the shear velocity of the surface layer. Therefore, the period of the second mode will be about 10 sec. To remove this, a digital filtering by the following moving average was applied to the records.

$$y_n = x_n + \frac{1}{2}(x_{n+2} + x_{n+1} + x_{n-1} + x_{n-2})$$

Since the coefficients are symmetric, no phase shifts are produced by this operation at any frequency. The amplitude response is given by

$$Y(\omega) = 1 + \cos(\omega\Delta t) + \cos(2\omega\Delta t)$$

Here Δt is the time interval between the sample points. If we take 3 sec for Δt , we have a response curve (Fig. 5) which has the minimum response at about 10 sec.

The smoothed seismograms for the two earthquakes are shown in Figure 6 with the original for earthquake July 64. The wave forms in the period range of 30 sec to more than 60 sec can now be used in the comparison. Figure 6 shows that the crest-to-crest correspondence is excel-

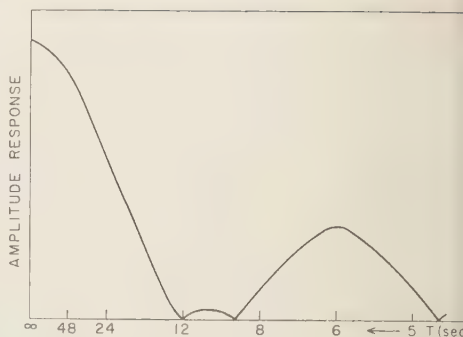


Fig. 5. Response curve of the digital filtering.

lent among these earthquakes, and therefore the fault motions for earthquakes July 6B and August 31 have the same direction in the lateral component as earthquake July 64. Here, again we see the significance of the fault plane solution. The fact that all the fault motions of these major earthquakes ($M = 6.0$ to 7.1) which occurred in a half-year period were right-hand strike slip tells us that the forces generating earthquakes are not very complicated and are uniform to a certain extent in time and space.

Is the source a single couple or a pair of couples? This question is very important and has not yet been satisfactorily resolved [*Byerly and Stauder*, 1957]. The single couple is used in North America and in Holland, and it seems to be the most successful in the Soviet Union. On the other hand, *Honda* [1957] showed many facts about Japanese earthquakes which favor his hypothesis of a pair of couples. *Ritsema* [1957] concluded, after a laborious study of the polarization of S waves of earthquakes in southeastern Asia, that the single couple is most likely the origin of the earthquakes in that region.

As is well known, the two models each give an identical quadrant distribution of P waves. The distribution of S waves, on the other hand, differs for the two cases. However, to find the first motion in S waves is very difficult.

The successful use of long-period Love waves for the study of earthquake mechanism suggest that this approach might be used to discriminate between the two types of sources. If the source is a single couple, the sense of the Love wave will be the same for any azimuthal direction. The sense of motion seen from the source is

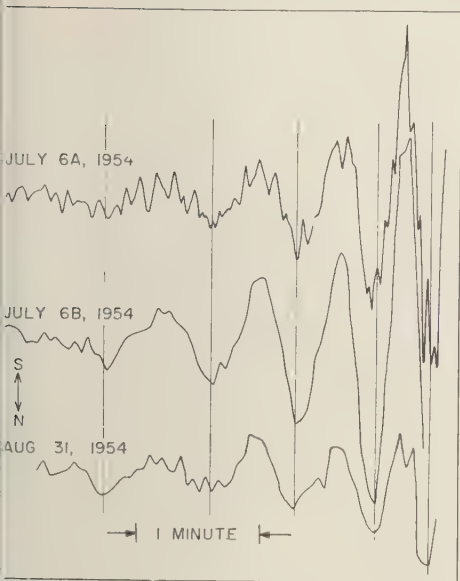


Fig. 6. Love waves from the Nevada shocks recorded at Palisades.

right hand for any direction, if the source is right-hand strike slip. On the other hand, if the source is a pair of couples, the distribution of the sense will be quadrant. If we measure the azimuthal angle to the station from the fault line, the sense of motion for the ranges of angles -45° to $+135^\circ$ and -45° to -135° will be right hand, and that for the ranges -45° to -45° and -135° to $+135^\circ$ will be left hand, in the case of right-hand strike slip.

Fortunately, the fault lines of the Kern County shocks are nearly perpendicular to those of the Nevada shocks (Table 1). We are able, therefore, to identify the type of source, by simply comparing the Love wave forms between the two series of shocks recorded at a single station. Let us take one shock of right-hand strike slip from the Kern County shocks and another from the Nevada shocks and compare the Love wave forms between the two shocks recorded at Palisades, for example. If the hypothesis of a single couple is true, both records must have the same sense. On the other hand, if the hypothesis of a pair of couples is true, they must have the opposite sense because the azimuthal angle measured from the fault line at

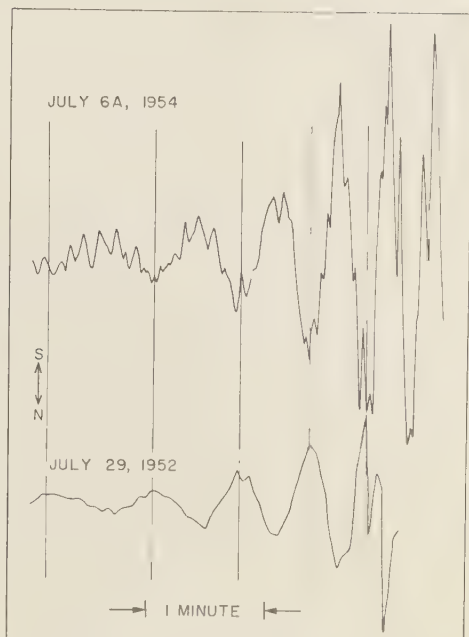


Fig. 7. Love waves from the Nevada shock of July 6A and from Kern County shock 141 recorded at Palisades.

the epicenter to Palisades is around 15° for the Kern County shocks and around 90° for the Nevada shocks. In this case, we may have to apply a correction to the wave forms with respect to the difference in epicentral distance and the structure of the earth's crust along the wave path.

At present, only the Palisades records are available for both series of shocks, but this station is the best because the difference in epicentral distance is small, around 100 miles. This difference cannot be neglected if shorter-period waves are used. As mentioned before, the difference Δx in the epicentral distance gives rise to a difference in the length of time from the maximum period crest to the minimum period crest which is roughly equal to $\Delta x(1/C_{\min} - 1/C_{\max})$, where C is the phase velocity. Reliable values of the Love wave phase velocity for 60 sec and 30 sec are obtained by the use of Brune, Nafe, and Oliver's [1960] method from the data of Kern County shocks, 4.03 km/sec for 30 sec and 4.38 km/sec for 60 sec. The difference in the length of time between the 60-sec crest and

the 30-sec crest will be about 3 sec for $\Delta x = 100$ miles. If we observe the wave forms of period longer than 30 sec this difference can be neglected and the direct comparison of seismograms without the correction can be allowed.

Also, the use of long-period waves may allow us to neglect the difference in sedimentary layers or topography between the wave path from Kern County to Palisades and that from Nevada to Palisades.

As mentioned before, the Love waves of the Nevada shocks recorded at Palisades are disturbed by shorter-period waves which may possibly be the second mode. The disturbance was more severe for the Kern County shocks, and long-period waves were relatively clear for only one of them, earthquake 141. The record of the Benioff long-period seismograph is smoothed by the operation described in the preceding section and is shown in Figure 7 with the record of the July 6A Nevada earthquake.

As is shown in the preceding sections, we know that the fault motion of shock 141 is right-hand strike slip, and that of shock July 6A is also right-hand strike slip. Despite the same sense in the slip motion at the source, Figure 7 clearly shows that the two shocks show the opposite sense at Palisades. The correspondence between the crests of earthquake 141 and the troughs of earthquake July 6A is excellent. If one arranges these wave forms to obtain a crest-to-crest correspondence, then, at the best arrangement, a difference between them of more than 15 sec is found in the time from the 60-sec crest to the 30-sec crest which cannot be accounted for by the difference in epicentral distance.

We cannot explain this fact by the hypothesis of a single couple, but we can explain it by Honda's hypothesis, in which a pair of couples generates the seismic waves. The azimuthal angle toward Palisades, measured from the fault line, is 20° for shock 141 and 97° for shock July 6A (Table 1); and, as mentioned before, when the senses of motions along the fault are the same for both earthquakes, their senses of motion at Palisades must be opposite if the earthquake source is a pair of couples.

Discussion. Though more studies must be made along this line before a firm conclusion is drawn, the result of this preliminary study of Love waves strongly favors the hypothesis of a pair of couples.

It is certainly not easy to explain the quadrant radiation pattern for Love waves from the generating force along a single fault plane. Honda explains it as follows: 'It may be considered to be natural to suppose that the vast region in the earth is subject to the large scale stresses corresponding to the force system of the type II (a pair of couples), that the earthquakes occur when some faults or sudden collapses are generated by the stress, and yet that the general features of the seismic waves and the deformation of the earth crust surrounding the focal region are governed by the large scale stresses.'

Another explanation may be suggested from the ultrasonic experiment by Press [1957] and Healy and Press [1959]. They found that shear waves from a secondary source distort the elementary pattern of the single couple and give rise to two nodal lines which lie at about 30° from the fault line. But, as our result is obtained from long-period Love waves, it might be difficult to assume a secondary source for them.

In a recent paper Knopoff and Gilbert [1959] calculated the radiation patterns for the first motions due to the sudden release of strain energy in the presence of a fault plane and showed the equivalent force systems that will yield the same radiation patterns in the absence of fault planes. They gave the solutions for eight independent elementary boundary conditions. One of the physically important solutions is for a displacement dislocation parallel to the fault. This has quadrant solutions for both P- and S-wave first motions and corresponds to a pair of couples in the absence of a fault. It seems highly probable, therefore, that an exact computation of the radiation pattern, taking into account the presence of a fault, also gives a quadrant solution for Love waves.

Acknowledgment. Frank Press suggested that the writer use surface waves having dispersed wave forms in the long-period range for the study of earthquake mechanism. The writer is most grateful for his original suggestion, as well as for his valuable advice during the course of the present study. The writer's thanks are also due Benoit Gutenberg, Hugo Benioff, and Charles F. Richter for their helpful advice and discussions. The writer's thanks also go to the seismologists at the Ottawa, Resolute Bay, Weston, and Palisades stations who furnished the data on which this study is based.

This work was partially supported by the IGY Interdisciplinary Research Program.

REFERENCES

- 38th, M., and C. F. Richter, Mechanism of the aftershocks of the Kern County, California, earthquake of 1952, *Bull. Seis. Soc. Am.*, **48**, 133-146, 1958.
- Brune, J. N., J. E. Nafe, and J. E. Oliver, A simplified method for the analysis and synthesis of dispersed wave trains, *J. Geophys. Research*, **65** (accepted for publication), 1960.
- Cyert, P., and W. V. Stauder, S. J., Motion at the source of an earthquake, *Publ. Dominion Obs.*, *Ottawa*, **20**, 255-261, 1957.
- Dorman, J., Numerical solutions for Love wave dispersion on a half-space with double surface layer, *Geophysics*, **24**, 12-29, 1959.
- Ewing, W. M., W. S. Jardetzky, and F. Press, *Elastic Waves in Layered Media*, McGraw-Hill, New York, 215 pp., 1957.
- Gutenberg, B., The first motion in longitudinal and transverse waves of the main shock and the direction of slip, Earthquakes in Kern County during 1952, *Calif. Dept. Nat. Resources, Div. Mines, Bull.* **171**, 165-170, 1955.
- Healy, J. H., and F. Press, Further model study of the radiation of elastic waves from a dipole source, *Bull. Seis. Soc. Am.*, **49**, 193-198, 1959.
- Hodgson, J. H., The null vector as a guide to regional tectonic patterns, *Publ. Dominion Obs.*, *Ottawa*, **20**, 369-384, 1957.
- Honda, H., The mechanism of the earthquakes, *Publ. Dominion Obs.*, *Ottawa*, **20**, 295-340, 1957.
- Knopoff, L., and F. Gilbert, First motions from seismic sources, *Seismic Scattering Project, Fourth Ann. Rept., sec. II, chap. X*, Inst. of Geophysics, Univ. of California, 1959.
- Press, F., Elastic wave radiation from faults in ultrasonic models, *Publ. Dominion Obs.*, *Ottawa*, **20**, 271-277, 1957.
- Richter, C. F., Foreshocks and aftershocks, Earthquakes in Kern County during 1952, *Calif. Dept. Nat. Resources, Div. Mines, Bull.* **171**, 177-198, 1955.
- Ritsema, A. R., On the focal mechanism of south-east Asian earthquakes, *Publ. Dominion Obs.*, *Ottawa*, **20**, 341-368, 1957.
- Romney, C., Seismic waves from the Dixie Valley-Fairview Peak earthquakes, *Bull. Seis. Soc. Am.*, **47**, 301-319, 1957.
- Tocher, D., Seismic velocities and structure in Northern California and Nevada, Doctoral thesis, Univ. of California, 1955.
- Tukey, J., Equalization and pulse shaping techniques applied to the determination of initial sense of Rayleigh waves, *Report on a Panel of Seismic Improvement, Appendix 9*, (Chairman, L. V. Berkner), 1959.

(Manuscript received August 24, 1959.)

A Comparison of Rank-Difference and Product-Moment Correlation of Precipitation Data

JAMES E. McDONALD AND CHRISTINE R. GREEN

*University of Arizona
Tucson, Arizona*

Abstract. A large sample (4650 cases) of comparisons of the Spearman rank-difference correlation coefficient r' and the Pearson product-moment correlation coefficient r is presented. The correlation data are 50-year records of winter and summer half-year precipitation totals for 220 stations, well distributed throughout the United States. The agreement between corresponding values of r' and r is very close; about 72 per cent of all cases differ by less than one-half the standard error of r , and about 94 per cent differ by less than one standard error of r . It is concluded that the distributions of r' and r are so close that, for most geophysical correlation applications r' is as adequate a statistic as r , and in certain cited cases is preferable to r ; hence r' deserves much wider use in climatology and hydrology than it has received in the past.

Introduction. In problems in climatology and hydrology where correlation techniques are needed it has become common practice to make almost sole use of the Pearson product-moment correlation coefficient. When it is desired to obtain regression equations for prediction purposes (for example, in certain rainfall-runoff studies), then the product-moment coefficient r is obtained almost as a by-product of the regression calculations. But in a much larger class of problems, one seeks only a reliable and objective measure of degree of association between two sets of data. For example, in pilot studies of climatological homogeneity of a region, one is chiefly interested in depicting the spatial pattern of correlation between one or more base stations and a large number of surrounding stations. In such problems, another type of correlation coefficient, the Spearman rank-difference correlation coefficient r' , deserves much wider use by geophysicists than it has received in the past. A particularly strong argument in favor of use of r' is found in its non-parametric nature, which makes it conceptually preferable to r for correlating non-normally distributed data, of which most hydrologic and climatic data are examples. But, in addition, ranking procedures are inherently faster than the procedures required to compute r , at least for sample sizes of less than 40 [Snedecor, 1946], particularly when manual methods rather

than electronic computer methods are to be used.

When r' is to be used merely to detect the existence of correlation between two sets of variates (that is, when one seeks only to answer the question of whether a sample value of r' differs from zero by an amount too great to be attributed simply to chance), then reference may be made to published significance tables [for example, Dixon and Massey, 1951]. But equally frequently, or perhaps even more frequently, the investigator wishes to draw conclusions from the magnitude of the coefficient itself (as in the aforementioned example of homogeneity studies), and here an obstacle is encountered. The sampling distribution of r' for parent bivariate populations of non-zero correlation has never been theoretically deduced. The geophysicist who seeks advice on this point in the literature of statistics will find that some statisticians suggest that the familiar standard error σ_r for the product-moment correlation coefficient will be quite trustworthy if applied to r' , whereas others say that sampling errors of r' simply cannot be assessed pending further theoretical developments for the non-zero case.

Because theoretical knowledge of the sampling distribution of such a statistic as r' is lacking, we must resort to empirical determinations. One published effort of this kind that has come to our attention is a very limited sampling experi-

ment described by *Snedecor* [1946, p. 166], involving only 10 trial samples. Its results indicate encouraging agreement between the distributions of r and of r' . The only other such experiment of which we are aware was done by *McDonald* [1957a], using only 14 precipitation correlations. In the latter study, 11 of the 14 cases gave r' values falling within one standard error of r , which is also fairly encouraging evidence that r and r' have closely similar distributions; but so small a sample is far from conclusive.

In the course of an extensive precipitation-correlation analysis that we have recently carried out on an automatic computer, we have had an excellent opportunity to secure a very much larger sample of comparison values of r and r' . The purpose of the present note is to summarize the results to show how very close an estimator of r the rank-difference coefficient r' is when applied to actual geophysical data.

Data and analysis. The present empirical comparisons were obtained as an incidental part of a study in which 50-year seasonal (winter and summer) precipitation records for 220 U. S. Weather Bureau stations were correlated, in groups of the order of 100 secondary stations each, against various base stations. Fifteen base stations, well distributed over the entire United States, yielded a total of 2325 summer and 2325 winter correlations, or a total sample of 4650 independent correlations. Each comprised time series of 50 seasonal-precipitation totals. These results will be reported elsewhere, so no listing of stations will be given here. 'Winter' was the period from November through April, 'summer' the balance of each year. All data were taken from published Weather Bureau records. The 50-year period was from 1906 to 1955. In a very small percentage of cases, missing monthly totals were estimated by the method described by *McDonald* [1957b]. Selection of the 15 base stations was done on the basis of double-mass analysis in order to be sure that all the base station records were homogeneous. The records of the secondary stations were not tested by double-mass methods; they were taken as published.

To prepare for the computation of the rank-difference correlation, the winter records and summer records of each of the 220 stations were

separately ranked by punchcard sorting techniques. Then, in the main part of the computing program, the computer not only performed the steps required to obtain the product-moment correlation, but concurrently took rank-differences and from these computed r' according to the equation

$$r' = 1 - 6 \sum_{i=1}^N (D_i^2) / N(N^2 - 1) \quad (1)$$

where D_i is the difference in ranks of the i th pair of variates and N is the total number of pairs, here 50 in every instance. The 'tied-ranks' problem, about which the statistics literature contains a variety of discussions, was simply ignored because it was found that treating the ties by the method specified in standard statistical references could not be expected to change numerical values of r' by more than about 1 per cent in the least favorable cases. The tied ranks were simply ranked in the *chronological* order in which they inevitably came through the ranking operation.

With so large an empirical distribution-sample as was involved in this study, it was indispensable to have some scheme by which the computer itself would do virtually all the comparison analysis, so the following simple method was employed. For each of the 4650 pairs of r and r' values, we calculated a quantity q defined according to

$$q = (r' - r) / \sigma_r \quad (2)$$

where σ_r is the conventional standard error of r given by

$$\sigma_r = (1 - r^2) / (N - 1)^{1/2} \quad (3)$$

Thus q measures the amount by which each r' exceeds its corresponding r , expressed in units of the standard error of r itself. As a final step in processing the data, the entire sample of 4650 correlations was sorted with respect to r , without regard to station or season, in order to group the results by r intervals of 0.20; and then within each r interval the result cards were ranked by value of r' to permit the final tallying, which is summarized below. This last procedure was employed to permit detection of any trends in the q distribution as r ranged from positive values near 1.0 (nearby station

TABLE 1. Percentage Distributions of q Values

r interval	q interval									Number of cases
	1.1 to 1.5	0.6 to 1.0	0.1 to 0.5	-0.4 to 0.0	-0.9 to -0.5	-1.4 to -1.0	-1.9 to -1.5	-2.4 to -2.0	-2.9 to -2.5	
.70 to .50	.89	2.1	4.1	9.3	24.8	32.0	17.6	6.2	3.1	97
.50 to .30	.69		7.2	20.8	31.3	19.1	10.6	6.6	1.0	303
.30 to .10	.49	0.6	6.0	26.9	37.2	19.4	8.5	1.5		862
.10 to -.09	.29	0.8	7.7	33.8	39.2	14.7	2.8	0.8		1402
-.09 to -.29	.09	0.8	9.4	39.0	40.6	8.2	1.5			1315
-.29 to -.49	0.9	0.9	8.6	42.8	38.9	8.4				581
-.49 to -.69	1.1	11.6	47.2	29.1	9.3					86
-.69 to -.89	1.0	0.7	8.0	34.0	38.2	13.5	3.9	1.2	0.3	4646

pairs) to the maximum negative values of almost -0.5 . This proved useful, for an interesting trend did appear, as will be noted below.

Results. In Table 1 are presented the percentage distribution of q values, grouped here in r intervals of 0.2 to facilitate examination. The merely four cases wherein r exceeded 0.90 are omitted, so the total sample comprises 1646 cases in Table 1. The number of cases falling within each r interval varied widely, since the original selection of stations was, of course, made without knowledge of how the r values would come out. That so many values of r lie near zero resulted from our extension of the correlation-field analyses out to distances often exceeding a thousand miles from the base stations. Even in the two least populous r intervals of Table 1, however, the subsamples are large enough to permit reliable conclusions to be drawn.

Discussion and conclusions. The results of principal interest in Table 1 are found in the bottom line of the table. For the entire sample, 72.2 per cent of the q values were of absolute magnitude equal to or less than 0.5; that is, this high percentage of r' values did not depart from their corresponding r values by more than one-half a standard error of r itself. And in 93.7 per cent of all cases r' lay within one standard error of r .

It seems justifiable to conclude that in almost all applications in which the geophysicist seeks a measure of correlation, r' will serve him just as well as r . Only in instances where regression

methods are to be employed or where it is established that the investigator deals with truly bivariate normal distributions could it be claimed that r will clearly be a better correlation statistic than r' . But the latter instances are almost unheard of in geophysical work, since non-normality is the rule, not the exception. Hence these results indicate that much wider use should be made of r' as an easily computed measure of correlation that is an impressively accurate predictor of r itself.

It will be noted that the maxima of the individual q distributions for the several class intervals of r in Table 1 exhibit a quite systematic trend, such that the mean absolute magnitude of r' tends to be somewhat less than that of r . Only for the class interval involving values of r between -0.09 and 0.09 is the q distribution symmetric about zero. Thus, it is generally true that r' is *slightly conservative* as an estimator of r . This property is also one that recommends use of r' , since, most of the time, one will not be misled into inferring an unduly large degree of association between variates when one employs r' rather than r .

The results described above fill in, at least on an empirical basis, the one significant gap in the arguments that can be advanced in support of viewing the Spearman rank-difference coefficient as the *preferred* measure of correlation in most geophysical application. These arguments have been discussed in some detail by McDonald [1957a], so here they need only be summarized very briefly:

1. The fact that r' is a distribution-free (non-parametric) statistic, plus the well-known fact that most rainfall, streamflow, and other geophysical data are not normally distributed, makes r' conceptually preferable to r . The vexing (but probably overemphasized) normality question simply disappears when an investigator measures correlation with r' rather than with r .

2. *Hotelling and Pabst* [1936] showed that r' has very high statistical efficiency even in the case of bivariate normal distributions. In the large-sample limit, and for normal bivariate populations characterized by zero correlation (the one theoretically soluble case), r' has an efficiency of 0.91. That is, in order to obtain rank correlations as sensitive (in the sense of equal sampling variance) as product-moment correlation coefficients, one must employ sample-sizes about 10 per cent larger for r' than for r , a consequence of information lost through converting from continuous variates to ranks. For non-normal cases, *Hotelling and Pabst* concluded that r' would have efficiency even higher than 0.91.

3. Not infrequently an investigator wishes to correlate quantities which are themselves nonlinearly related to the independent variables governing them. Relative humidities, vapor pressures, and kinetic energies of flow are familiar examples. When product-moment correlation of such quantities is carried out, there is reason to doubt whether some linearizing transformation of the quantities should precede correlation analysis. But, inasmuch as ranks are invariant under any kind of transformation that only stretches scales, this question entirely disappears when rank-correlation techniques are used, a point called to our attention by R. A. Bryson.

4. When correlation is to be done manually or by desk calculator, and when the sample size is less than about 40, r' can be computed more quickly than r . This advantage becomes particularly noticeable with sample sizes of only 20 or so, for then the ranking step proceeds very rapidly, since the squares of the rank-differences may be simply written from memory.

5. The large-sample comparison reported here strongly supports the conclusion that the sampling distribution of r' for the case of non-zero correlation (the case for which there remains no

theoretical solution) must be so nearly identical with that for r that, in most geophysical studies, entirely acceptable estimates of sampling error of r' can be obtained simply by using σ_r , the familiar standard error of r .

In the three most recently published works on the use of statistics in climatology and meteorology, there is so little information on r' that one is led to believe that this statistic has little to recommend it. *Conrad and Pollak* [1950] and *Panofsky and Brier* [1958] do not mention r' , and *Brooks and Carruthers* [1953, p. 235] go so far as to say that correlation by ranks is not likely to have much application in meteorology. We strongly disagree with the latter view, and recommend that r' be regarded as the preferred correlation statistic in all cases of moderate sample size, where (as is almost invariably true) the data are not known to be normally distributed, and where regression analysis is not the fundamental objective.

Acknowledgments. We wish to express our appreciation to Robert W. Mitchell of the University of Arizona Numerical Analysis Laboratory for his assistance in computational matters. The work reported here was supported by the Office of Naval Research under Contract NR 082-164.

REFERENCES

- Brooks, C. E. P., and N. Carruthers, *Handbook of Statistical Methods in Meteorology*, M. O. 538, Air Ministry, H. M. Stationery Off., London, 412 pp., 1953.
- Conrad, V., and L. W. Pollak, *Methods in Climatology*, 2d ed., Harvard Univ. Press, Cambridge, 459 pp., 1950.
- Dixon, W. J., and F. J. Massey, *Introduction to Statistical Analysis*, McGraw-Hill, New York, 370 pp., 1951.
- Hotelling, H., and M. R. Pabst, Rank correlation and tests of significance involving no assumptions of normality, *Ann. Math. Statist.*, 7, 29-43, 1936.
- McDonald, J. E., A critical evaluation of correlation methods in climatology and hydrology, *Sci. Rept. 4, Inst. Atm. Physics*, Univ. Arizona, 36 pp., 1957a.
- McDonald, J. E., A note on the precision of estimation of missing precipitation data, *Trans. Am. Geophys. Union*, 38, 657-661, 1957b.
- Panofsky, H. A., and G. W. Brier, *Some Applications of Statistics to Meteorology*, Pennsylvania State Univ., University Park, Pa., 224 pp., 1958.
- Snedecor, G. W., *Statistical Methods*, 4th ed., Iowa State Coll. Press, Ames, Iowa, 485 pp., 1946.

(Manuscript received August 29, 1959.)

Rb-Sr and K-A Ages of Rocks in Ontario and Northern Minnesota

L. T. ALDRICH AND G. W. WETHERILL

*Department of Terrestrial Magnetism, Carnegie Institution of Washington
Washington, D. C.*

Abstract. Analyses of micas from seven localities in Southern Ontario with previously published data broadly outline the area of the oldest rocks in the Canadian Shield. The ages as determined by both the Rb-Sr and K-A methods range from 2500 to 2650 million years.

Rocks having ages in excess of 2000 m y have been shown to exist on every continent. The outlines of the areas of exposure of these rocks are known only approximately. It is the purpose of this paper to present data, giving these ages, on additional samples collected in central Ontario from Kirkland Lake to Kenora. These data with those already published help to further define the boundaries of this ancient shield area as given by presently exposed rocks. The decay constants and equations used to calculate the ages have been discussed earlier

[Aldrich and Wetherill, 1958]. The isotopic abundances of the parent elements used are given in table 1 and the decay constants in table 2. The analytical procedures are based on the method of isotope dilution, and no changes have been made, other than to narrow the range of errors wherever possible, from procedures previously published [Wetherill and others, 1956; Aldrich and others, 1956]. The details of the sources and petrographic examinations of the rocks used for this study are given in the Appendix.

TABLE 1. Analytical Data

Location*	Mineral	K ⁴⁰ , † ppm	Radio- genic A ⁴⁰ , ppm	A ⁴⁰ Radio- genic, per cent	Rb ⁸⁷ , ‡ ppm	Radio- genic Sr ⁸⁷ ppm	Sr ⁸⁷ Radio- genic, per cent
Inclusion in vermillion granite South of International Falls, Minnesota	Biotite	5.55	1.89	92	105.	3.93	65
Granite, east of Kenora, Ontario	Biotite	9.24	2.81	98	133.6	4.81	75
Granite gneiss, Sioux Lookout, Ontario	Biotite	7.39	2.13	99	129.	4.01	66
Granite, west of Keefer Twp., Ontario	Biotite	6.31	1.78	98	81.9	2.91	58
Inclusion, Round Lake batholith, Otto. Twp., Ontario	Biotite	7.44	2.31	98	106.	3.70	32
Granite, Round Lake batholith	Biotite	9.60	3.07	99	83.4	3.01	56
Biotite-rich rock, Cobalt, Ontario	Biotite	7.80	1.61	98	124.2	3.90	73

* Detailed locations given in Appendix.

† K⁴⁰ = 1.22×10^{-4} g/g K.

‡ Rb⁸⁷ = 0.283 g/g Rb.

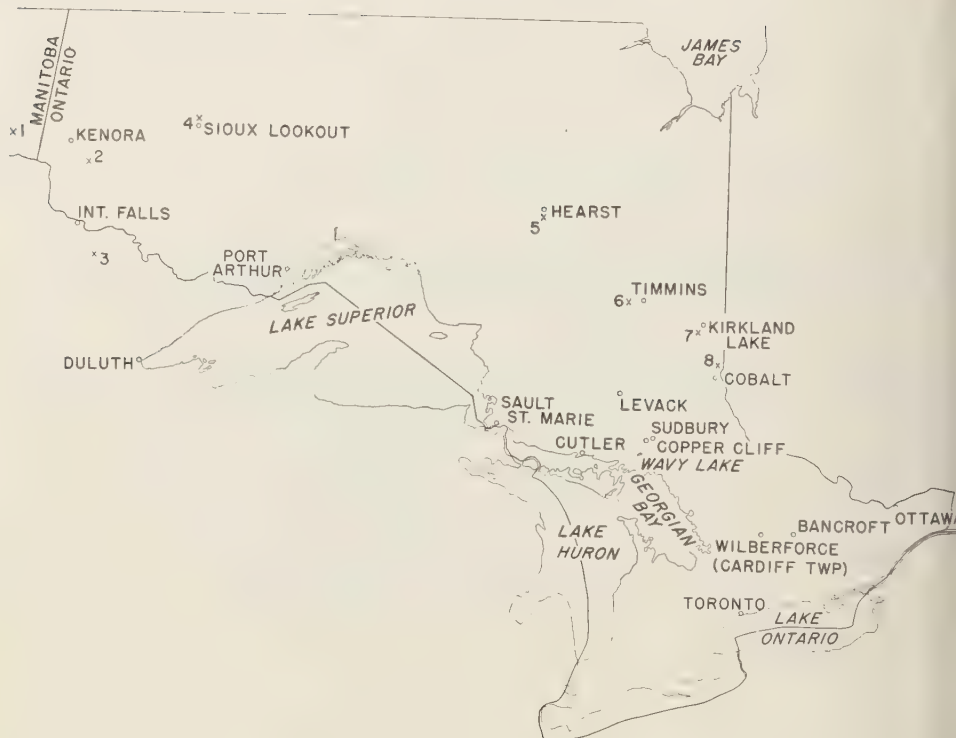


Fig. 1. Map showing locations of samples analyzed.

The results of the analytical measurements made are given in Table 1. The absolute analytical errors are believed to be less than 5 per cent and are regularly reproduced within a range of 1 per cent. The mean deviation of the isotopic abundances measured for the analyses is $\frac{1}{2}$ per cent or less. One feature of the analyses themselves is the striking uniformity of both the Rb and Sr content of the biotites.

The ages derived from these measurements and others from this area reported earlier are given in Table 2. These data indicate that rocks of this age extend over the entire width of the Province of Ontario. The data of *Wilson and others* [1956] in western Quebec and of *Gast and others* [1958] in eastern Manitoba extend the margins of the ancient shield in the east-west direction. The measurements of *Goldich and others* [1957, and in preparation] fill in the region to the south. Figure 1 shows the locations

of the samples analyzed at the Carnegie Institution. The numbers on the map refer to the numbers of Tables 1 and 2.

Rocks younger than 2600 m y old have been found on all the margins of this area, and some have even been measured inside the area. Many of the rocks 2600 m y old have been metamorphosed, and thus 2600 m y is a lower limit to the age. The samples at Sioux Lookout and a Cobalt are included because they show an age pattern which could be the result of the metamorphism of rocks 2600 m y old.

APPENDIX

Sample locations and descriptions. Sample locations are given from the best available maps. The descriptions are from thin section examined by G. L. Davis of the Geophysical Laboratory, Carnegie Institution of Washington. The sample numbers refer to location

TABLE 2. Summary of Age Determinations

Sample	Mineral	Age, m y*	
		Rb-Sr	K-A
1. Silverleaf Mine, pegmatite†	Lepidotite	2680	2160
2. Granite, east of Kenora	Biotite	2550	2490
3. Inclusion in vermilion granite	Biotite	2640	2670
4. Granite gneiss, Sioux Lookout	Biotite	2200	2420
5. Pegmatite, Hearst, Ontario†	Lepidotite	2600	2540
6. Granite, Keefer Township	Biotite	2510	2390
7. Inclusion, Round Lake batholith	Biotite	2470	2520
8. Granite, Round Lake batholith	Biotite	2550	2570
9. Biotite-rich rock, Cobalt, Ontario	Biotite	2230	1990

* Ages based on: $\lambda(\text{Rb}^{87}) = 1.39 \times 10^{-11} \text{ yr}^{-1}$

$\lambda_e(\text{K}^{40}) = 5.80 \times 10^{-11} \text{ yr}^{-1}$

$\lambda_i(\text{K}^{40}) = 5.30 \times 10^{-10} \text{ yr}^{-1}$

† See Aldrich and others, [1958] for analytical data and exact locations.

shown on the map (Fig. 1). Samples 1 and 5 have been described previously.

2. Vermilion granite inclusion collected from a road cut on U. S. Highway 53 2 miles west of the Kabetogama Lake Road southeast of International Falls, Minnesota. The sampling site was at $93^{\circ}5'W$, $48^{\circ}24'N$. Biotite fresh.

3. Granite. Road cut in northeast corner of Kirkup Township on Highway 70, 1.6 miles south of junction with Highway 17, $49^{\circ}40'N$, $94^{\circ}15'W$. Gray granite with fine-grained, fresh biotite, magnetite, zoisite, epidote, and carbonate.

4. Gneiss. From an island in Pelican Lake, north-northeast of Sioux Lookout, $50^{\circ}8'N$, $91^{\circ}57'W$. Located from map of Hurst [1932]. Part of granite gneiss on the northern boundary of the outcrops of Hurst's Timiskaming and Keewatin sediments. Thin section shows evidence of alignment of fresh biotite. Accessories also present are mildly chloritized, flake muscovite, zoisite, and magnetite.

6. Granite. Keefer Township from a road cut

approximately one mile west of boundary between Keefer and Denton Townships and about one mile south of the center line through Keefer Township, $48^{\circ}18'N$, $81^{\circ}46'W$ on map of Harding and Berry (1938). Collected under the direction of S. Ferguson of the Ontario Department of Mines. Granite altered with chlorite, fresh biotite, zoned plagioclase, and sericitized potassium feldspar.

7. Biotite-rich inclusion in the syenite phase of Round Lake batholith from a road cut in Otto Township, concession 5, lot 3, $48^{\circ}02.5'N$, $80^{\circ}01'W$. Biotite fairly fresh in this metamorphosed sample.

8. Granite. Round Lake batholith, collected from a road cut in the east-central part of Savard Township, concession 5, lot 1, $47^{\circ}54'N$, $80^{\circ}02'W$. Two-feldspar granite with dull-green biotite, muscovite, epidote zoisite amphibole, and slightly sericitized potassium feldspar. Samples 7 and 8 collected under the direction of W. S. Savage of the Ontario Department of Mines.

9. Biotite-rich rock collected from the Haileyburian formation, which is post-Timiskaming and pre-Cobalt in sequence. It is a metamorphic rock with quite fresh biotite. Collected from concession 6, lot 5, of Coleman Township ($47^{\circ}23'N$, $79^{\circ}41'W$) under the direction of Robert Thompson of the Ontario Department of Mines.

REFERENCES

- Aldrich, L. T., G. L. Davis, G. R. Tilton, and G. W. Wetherill, Radioactive ages of minerals from the Brown Derby Mine and the Quartz Creek granite near Gunnison, Colorado, *J. Geophys. Research*, **61**, 215-232, 1956.
- Aldrich, L. T., and G. W. Wetherill, Geochronology by radioactive decay, *Ann. Rev. Nuclear Sci.*, **8**, 257-298, 1958.
- Aldrich, L. T., G. W. Wetherill, G. L. Davis, and G. R. Tilton, Radioactive ages of micas from granitic rocks by Rb-Sr and K-A methods, *Trans. Am. Geophys. Union*, **39**, 1124-1134, 1958.
- Gast, P. W., J. L. Kulp, and L. E. Long, Absolute age of early Precambrian rocks in the Bighorn Basin of Wyoming and Montana, and southeastern Manitoba, *Trans. Am. Geophys. Union*, **39**, 322-334, 1958.
- Goldich, S. S., H. Baadsgaard, and A. O. Nier, Investigations in A^{40}/K^{40} dating, *Trans. Am. Geophys. Union*, **38**, 547-551, 1957.
- Harding, W. D., and L. G. Berry, Annual Report,

- Ontario Dept. of Mines, Vol. XXVII, Part 4, 1938.
- Hurst, M. E., Geology of Sioux Lookout Area, Annual Report, Ontario Dept. of Mines, Vol. XLI, Part 6, 1932.
- Wetherill, G. W., G. R. Tilton, G. L. Davis, and L. T. Aldrich, New determinations of the age of the Bob Ingersoll pegmatite, Keystone, Saskatchewan, *Geochim. et Cosmochim. Acta*, 9, 292-297, 1956.
- Wilson, J. T., R. D. Russell, and R. M. Farquhar, Economic significance of basement subdivision and structures in Canada, *Trans. Can. Mining Inst.*, 59, 310-318, 1956.

(Manuscript received October 2, 1959.)

Isotopic Evidence on the Origin and Age of the Blind River Uranium Deposits

J. A. MAIR,¹ A. D. MAYNES,² J. E. PATCHETT, AND R. D. RUSSELL³

*Department of Physics, University of Toronto
Toronto, Ontario, Canada*

Abstract. Isotopic analyses of lead extracted from a variety of minerals from Blind River, Ontario, are reported. The detrital minerals monazite and zircon both give lead-ratio ages of 2500 million years. The uraninite ore gives a lead-ratio age of 1700 m.y. Other isotopic evidence is quoted to suggest that the age of the sediment in which the uranium is found may also be approximately 1700 m.y. or older. The lead found in pyrite, pyrrhotite, sericite, and feldspar has anomalous isotopic ratios which can be explained by the hypothesis that they received additions of radiogenic lead from the uraninites (presumed to be 1700 m.y. old) 1200 to 1300 m.y. ago. In any case the age of these minerals, in the sense of time of last chemical alteration, is not greater than 1450 ± 150 m.y. All our measurements can be interpreted without assuming a major period of mineralization more recent than 1000 m.y. ago, although we are unable to rule out such a possibility from our evidence.

INTRODUCTION

For several years the Geophysics Laboratory at the University of Toronto has been carrying out extensive investigations of the natural variations of the isotopes of common ore-lead minerals. We have now modified the Metropolitan Vickers type M.S.2 mass spectrometer operated at the laboratory, by the addition of a solid source and an electron multiplier. Adequate sensitivity can be obtained from samples of lead as small as $1 \mu\text{g}$ with the modified instrument. So we have developed a new method for the extraction of lead from silicate mineral samples where the concentration is in the part-per-million range [Maynes and McBryde, 1957]. Because of our interest in the age relationships of the Keewatin and Grenville provinces of Ontario, and the complex rocks along their common boundary, it was decided to use the facilities developed in an investigation of this problem. Specifically, we have determined the isotope ratios for some samples of lead from the Blind River mining area.

The Blind River uranium deposits have been described by Roscoe [1957], among others, and much of the following description is abstracted from his paper. The deposits, which are sheet-like in character and classified as pyritic, quartz-pebbled conglomerates, are in gently dipping beds at or near the base of the Huronian sediments. The Huronian rocks, which overlie unconformably a granite and greenstone basement, include quartzite, argillite, various types of conglomerates, and minor limestone. They have been folded to form a syncline plunging gently toward the west and are weakly metamorphosed.

The Proterozoic sedimentary rocks of the north shore of Lake Huron have been divided into a lower, Bruce, series and an upper, Cobalt, series, which Collins [1925] believed were separated by an unconformity. Collins divided the Bruce series (bottom to top) into Mississagi formation (mainly quartzite), the Bruce conglomerate, the Espanola formation, and the Serpent formation. The Mississagi formation is in turn divided by Roscoe into four formations, of which the lowermost is the Matinenda. In two places within the Quirke Lake trough the Matinenda becomes abnormally thick and forms 'valley' structures, and it is in these two 'valleys' that most of the uranium ore occurs. One of these structures extends southwest from the Algoma Quirke mine and the other extends northwest from the Algoma Nordic mine. The ore is

¹ Present address: Imperial Oil Ltd. Research Laboratory, 300-9th Avenue West, Calgary, Alberta.

² Present address: Department of Geological Sciences, California Institute of Technology, Pasadena, California.

³ Present address: Department of Physics, University of British Columbia, Vancouver 8, B. C.

known to extend 5 and 4 miles from these mines, respectively.

The thickness of the ore zone is generally about 10 feet, but locally reaches 32 feet. The ore deposits consist of interlayered beds, 1 foot to 3 feet thick, of quartz-pebbled conglomerate, conglomerate quartzite, and pebble-free quartzite.

There are two principal theories of the origin of the uranium mineralization. In the first the uranium minerals are assumed to have had a placer origin, and in the second they are assumed to have precipitated from hydrothermal solutions. Neither of the mechanisms seems compatible with all the observations.

The strongest evidence for the placer hypothesis is the strong sedimentary control on both regional and local scale. The richest ore sections occur in the thickest portions of the Matinenda formation, especially where the conglomerates consist of closely packed pebbles. Concentrations of uranium minerals are also noted in streaks and in such features as cross-bedding. A weakness of this hypothesis is the instability of uraninite under weathering conditions. *Davidson* [1957] enlarges on this aspect and is convinced that uraninite could not survive the chemical and physical processes to become an important constituent of the conglomerates. Thus those who support a placer origin for the uraninite resort to special environmental conditions to explain the persistence of uraninite.

The strongest evidence for the hydrothermal hypothesis is the detailed mineral textures which are similar to those of some hydrothermal occurrences. (Proponents of the placer hypothesis consider these to have been caused by pseudo-hydrothermal recrystallization.) The quantities of associated sulphide minerals are easily explained, and the types of sulphides are compatible with other known uranium hydrothermal occurrences. The weaknesses of the hydrothermal theory are largely that there are no obvious solution channels, no wall rock alteration, and no easy means of explaining the disseminated mineralization. It is difficult to explain by a hydrothermal theory why discontinuous conglomerate beds are ore bearing when the interbedded quartzites are barren. It is probable that we do not now possess sufficient

information to enable us to make a firm decision as to the mode of formation.

AGE RELATIONSHIPS

The available age data obtained by the lead-alpha, lead isotope, rubidium-strontium, and potassium-argon methods on rocks directly or indirectly related to the Mississagi or its source rock are in substantial disagreement. *Aldrich and Wetherill* [1958] have published rubidium and potassium ages for the Cutler batholith and its pegmatites, which, according to *Collins* [1925], intrude the Mississagi. The ages for the granite by the Rb-Sr and K-A methods agree, and a mean value of 1330 m y is obtained as a lower limit to the formation of the Mississagi. Unfortunately, pegmatites cutting both the Cutler batholith and neighboring Archean-type sediments give discordant and older ages (1350 to 1800 m y). Thus, if the Mississagi is intruded by the Cutler granite, the sediment must certainly be older than 1350 m y and possibly older than 1750 m y.

Ginn [1959] questions the age relationships of the Cutler granite and the Mississagi. Near the town of Cutler, Ontario, the two are separated by at least half a mile, and this interval is a major fault zone. The brittle and glassy character of the Mississagi quartzite in this area, attributed to contact metamorphism by *Collins* [1925], could have been caused by other means. The actual evidence that was seen farther to the south on John Island is not described by *Collins*, but a fault zone is thought to occur there as well. Thus the field relationships of the granite and sediment are not clearly defined.

Recent field studies by *Ginn* [1959] show that in Lorne and Nairn townships the Middle Mississagi sediments are cut by basic dykes which metamorphose the sediments. *Ozima* [1958] found potassium-argon ages in the range of 1400 to 1600 m y for the biotite and sericite formed by the metamorphism. Moreover, *Ozima* obtained an age of 1720 m y for mica, separated from a granite boulder within the Bruce Conglomerate which overlies the Mississagi.

Published analyses by *Wanless and Traill* [1956] of galena and uraninites are listed in Table 1. The age of 1300 m y for sample 4 is concordant. An age of about 600 m y was found for sample 5, but the Pb^{204} and Pb^{208} abundances

TABLE 1. Analyses by *Wanless and Traill* [1956] of Galena and Uraninite

Sample	204	Atom, per cent		208	Pb, per cent	U, per cent	Th, per cent
		206	207				
Galena	1.42	24.30	22.67	51.61			
Galena	1.06	36.04	15.99	46.92			
Galena	0.76	45.21	14.14	39.89			
Uraninite	0.10	86.15	8.19	5.59	10.7	43.8	3.6
Uraninite	0.33	81.42	9.82	8.43	5.2	48.3	4.5

Ages Determined from Uraninites, Corrected Relative to Galena No. 1

Sample	Pb ²⁰⁶ /U ²³⁸	Pb ²⁰⁷ /U ²³⁵	Pb ²⁰⁷ /Pb ²⁰⁶
4. Uraninite	1400 my	1300 my	1170 my
5. Uraninite	590 my	600 my	630 my

is inconsistent and therefore the age is in some doubt. If the age of the mica from the granite boulder is the actual age of the granite, then the Huronian sedimentation at Blind River is to be younger than 1750 m.y. However, our evidence does not allow us to exclude the possibility that the mica has been recrystallized in the sediment or that there has been a loss of argon.

Isotopic analyses of the uranium ore by Stieff and Stern have been interpreted by them to indicate a Paleozoic age, a conclusion at variance with those reached in this paper. Their interpretation is presented only in the form of a brief abstract, and we shall not discuss it until a more detailed account is available.

LEAD ISOTOPE ANALYSES

In this study, analyses of the lead isotope ratios were determined for detrital monazite, zircon and feldspar, uraninite, brannerite, pyrochlore, sericite, and pyrite. The minerals were separated from the ore samples by J. E. Hutchett and the lead extracted from the minerals by A. D. Maynes, using modifications of the technique described earlier [Maynes and McBryde, 1957]. All results include a correction for electron multiplier discrimination that was obtained by repeating analyses of test samples with a simple Faraday cup rather than with the multiplier. The following results were obtained. *Monazite and zircon.* The monazite grains are yellow to orange to amber crystal frag-

ments, often well rounded. Although monazite is easily recognized by its subresinous to vitreous luster, it was difficult to distinguish some monazite grains from zircon and xenotime, and these probably contaminate the sample by 2 to 3 per cent. Typical grains of monazite included in this sample produced an X-ray diffraction pattern with no extra lines.

Monazite is a true detrital mineral and there is little possibility of secondary overgrowths. Therefore its age should give a close approximation to the age of the detrital minerals.

The Geological Survey of Canada has in preparation a mineralogical study of the ore conglomerates from the Blind River mines, and, since large bulk samples were used, zircon concentrates were obtained in sufficient quantities for micro-lead analyses. Zircon is one of the very rare minerals, only 1 to 2 grams being recovered from 70 pounds of the conglomerate. S. M. Roscoe of the Mineral Deposits Division made available a zircon concentrate from the lower conglomerate horizon at the Algou Quirke Mine bearing the number G.S.C. 462. Amber and purplish crystal fragments were hand picked for the analysis, and the resulting sample was contaminated with perhaps 2 to 3 per cent xenotime and monazite. Where the entire crystal is seen it is generally well rounded.

The isotopic analyses for lead from the zircon and monazite are shown in Table 2. Both analyses are corrected for the presence of a blank.

Although no uranium or thorium analyses

TABLE 2. Analyses of Detrital Monazite and Zircon

Sample	Blank Lead, gm	Lead Content, ppm	Atom, per cent*				Lead-Ratio Age, m y
	Total Lead, gm		204	206	207	208	
Monazite	0.017	3100	0.0259 (20%)	13.12 (0.6%)	2.48 (2%)	84.37 (0.3%)	2550
Zircon	0.093	190	0.0555 (25%)	70.48 (0.5%)	11.56 (1%)	17.90 (1%)	2450

* The figures in parentheses represent estimated uncertainties in view of the particular mass spectrometric analysis and correction for blank.

were carried out, and therefore no test for concordancy could be applied, there is good reason to trust lead-ratio ages, particularly for samples of such great age. Lead-ratio ages in this age range have never been shown to be in serious error. Of concern is the rather large correction required for the presence of nonradiogenic lead contamination, particularly in the case of the monazite. The order of magnitude of the uncertainty involved can be estimated by recalculating the ages without making any correction for common lead. If this is done, the zircon age is increased by 4 per cent and the monazite age by 9 per cent. These are certainly overestimates of the uncertainties involved; therefore the writers consider that the ages of these detrital minerals

are in agreement at 2500 m y and that the uncertainty in this value is not greater than 100 m y.

Uraninite and brannerite. Brannerite and uraninite are the major ore minerals. The former may be the more important, but the data on its distribution and abundance are not yet available. Brannerite is a metamict mineral, and identification by X-ray diffraction can only be made after prolonged heating at red heat to restore the crystal structure. Brannerite was originally reported as a placer mineral, but it has since been reported in metamorphic, igneous and hydrothermal environments. The Blind River brannerite is intimately associated with anatase, and the leaching of uranium has been suggested to account for the excess TiO_2 .

Lead from two uraninite samples was studied. The first sample (A), 1.93 mg in size, consisted of crystals and crystal fragments with sharp edges and corners showing no sign of abrasion or solution. The crystals are cubes, more or less distorted. The purity of the sample is believed to be greater than 99 per cent. The second sample (B) consisted of fresh, unaltered anhedral fragments that are weakly magnetic. The X-ray diffraction pattern consists of weak pyrrhotite lines superimposed on the UO_2 and galena patterns. The sample weighed 2.33 mg.

Two brannerites were studied. The first sample, A, reddish brown with subresinous luster produced strong anatase lines in the unheated diffraction pattern. The second sample B, dark brown to black, yielded a weaker anatase pattern and may be less altered than brannerite A (Table 3).

The ages of the uraninites are in agreement within their uncertainties. Sample A, which was

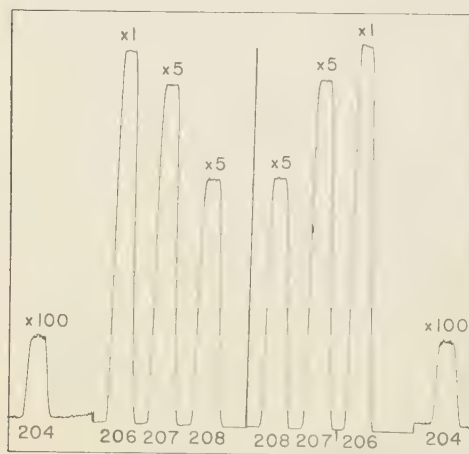


Fig. 1. Unretouched photostat of spectrogram made with 10 μg of pyrite lead. The Pb^{204} peak represents only 0.16 per cent of the total sample or 0.016 μg .

TABLE 3. Analyses of Uraninite and Brannerite

Sample	Blank Lead, gm	Lead Content, ppm	Atom, per cent				Lead-Ratio Age, m y
	Total Lead, gm		204	206	207	208	
Uraninite A	0.0048	85,000	0.0097 (20%)	87.39 (0.2%)	8.98 (0.2%)	3.60 (0.5%)	1680
Uraninite B	0.0070	63,000	0.0123 (20%)	86.49 (0.2%)	9.23 (0.3%)	4.28 (0.8%)	1740
Brannerite A	0.043	33,000	0.0251 (2%)	84.96 (0.3%)	10.21 (0.3%)	4.80 (0.2%)	1870
Brannerite B	0.060	39,000	0.035 (25%)	85.48 (0.3%)	10.80 (1.8%)	3.35 (1%)	2000

The figures in parentheses represent estimated uncertainties in view of the particular mass spectrometric analysis and correction for blank.

The figures in parentheses represent estimated uncertainties in the mass spectrometer analyses alone. Although very large, the uncertainty due to blank correction does not affect the calculated age significantly.

pure, is considered to have given the better age because uraninite B was contaminated with pyrrhotite and a common lead correction was required. The pyrrhotite contains highly radioactive lead, rich in Pb^{207} , and the listed age for uraninite may be slightly too great because of this impurity. In consideration of these ages, it seems likely that the age of the uraninite actually approximates that of the sediments or is younger.

The age values of the brannerite are less acceptable. The fact that the brannerite is altered will have some effect on the lead-ratio age, even though this age is less sensitive to the effects of chemical alteration than most others. We do not know when the alteration took place or presume that it was during the early history of the mineral. Uranium has probably been lost, and if there has been little loss of lead the apparent age is too high. On the other hand, if more lead than uranium has been lost the apparent age is too low. Only gross losses could introduce an error much in excess of 200 m y. Therefore it is impossible to say from this data when the brannerite formed, but an age as great as that found for the monazite and zircon seems highly unlikely.

Common Leads. Apart from those leads that are relatively rich in uranium and thorium, lead also found in minerals containing amounts of uranium and thorium too small to affect the isotope ratios. Such leads are here referred to as common leads.

The common leads found at Blind River are *anomalous* leads. Such leads, which are characterized by the addition of substantial quantities of radiogenic lead at a single period in the lead's history, have been shown to be particularly useful in studying the history of ore deposits [Farquhar and Russell, 1957; Russell, and others, 1957; Stanton and Russell, 1959]. Plots of such leads (Pb^{207}/Pb^{204} vs. Pb^{206}/Pb^{204}) from a single locality may be fitted by a straight line, the slope of which is related to the history of the lead mineral. The value of this approach is that it is independent of theories of the primary origin of lead ores.

Specifically, during an interval of time beginning at t_1 and ending at t_2 , a sample of natural uranium will have produced lead isotopes in the ratio

$$Pb^{207}/Pb^{206} = \frac{1}{\alpha} \frac{\epsilon^{\lambda' t_2} - \epsilon^{\lambda' t_1}}{\epsilon^{\lambda t_2} - \epsilon^{\lambda t_1}} \quad (1)$$

where

$$\lambda = 0.1537 \times 10^{-9} yr^{-1}$$

$$\lambda' = 0.9722 \times 10^{-9} yr^{-1}$$

$$\alpha = 137.8$$

For this study, isotope ratios were measured for lead extracted from galena, pyrrhotite, pyrite, sericite, and feldspar. The analyses are given in Table 4.

The feldspar occurred in large rounded clas-

TABLE 4. Isotopic Analyses of Common Leads

Sample	Blank, gm	Lead Content, ppm	Atom, per cent*			
	Total Lead, gm		204	206	207	208
Pyrrhotite	0.0051	37	0.067 (10%)	79.58 (0.5%)	12.54 (0.5%)	7.82 (0.5%)
Pyrite	0	...	0.1639 (0.3%)	76.08 (0.2%)	13.79 (0.2%)	9.96 (0.2%)
Sericite	0.1†	...	0.145 (2%)	78.63 (0.2%)	11.58 (0.6%)	9.62 (1%)
Feldspar residue	0.157	13	0.629 (2%)	56.04 (1%)	17.9 (1%)	25.45 (2%)
Feldspar wash	0.49		0.191 (20%)	86.51 (2%)	6.58 (15%)	6.77 (15%)
Galena Quirke Mine, University of Toronto analysis by lead tetra- methyl method (T-988)		...	0.567 (1%)	45.25 (0.3%)	13.75 (0.3%)	40.43 (0.3%)

* The figures in parentheses represent estimated uncertainties in view of the particular mass spectrometer analysis and correction for blank.

† Estimated. Possible error in this figure not considered in estimating uncertainties.

tic grains in a conglomerate sample supplied by D. H. Gorman from the Pronto Mine. Since these are very unlikely to be a second-cycle component in a sediment, their age should represent that for some of the pegmatitic source rock.

The feldspar grains were obtained by cutting out of the rock the portion containing the feldspar fragment and then grinding away the surrounding material. Several individual grains so obtained varied in weight from 0.2 to 1.0 gram. Pyrite in very minute crystals was an obvious impurity in most of the grains and appeared to be localized along fractures. Grinding did not remove all the pyrite, as new grains were surfaced as others were removed. To minimize this contamination the one feldspar grain that was examined was leached with aqua regia for 12 hours. The leaching extracted nearly 40 per cent of the lead, although only 1 per cent of the total sample was dissolved. The analyses of the aqua regia solution and the feldspar residue are reported separately in Table 4. Although less radiogenic than the wash, the feldspar lead is still remarkably anomalous.

When these points are plotted (Pb^{207}/Pb^{204}) vs. Pb^{206}/Pb^{204}) with the galena analyses of Wanless and Trail [1956] (Table 1), the com-

bined group fits rather imperfectly to a straight line of slope giving $Pb^{207}/Pb^{206} = 0.15$. The certainty in this value is not greater than per cent.

From equation (1) it is possible to choose a wide range of values of t_1 and t_2 that give a radiogenic lead ratio of 0.15. One obvious possibility is that the accumulation of radiogenic lead that was subsequently added to the anomalous leads began 1700 m.y. ago, at time of the uraninite formation. Setting $t_1 = 1700$ m.y. results in a value for t_2 of approximately 1200 m.y., which is the time when radiogenic lead was added to the minerals in question. There is ample evidence of sulphur mineralization at this time in neighboring regions; for example, at Sudbury [Russell and others, 1954] and at Theano Point [Collins and others, 1954]. Even at Cobalt, Ontario, anomalous leads (Russell and Farquhar, unpublished data) have a radiogenic component very similar to that at Blind River, a fact which supports the field interpretations of many geologists, including Joubin [1954], that the pyrites at Blind River are indeed similar to those in sediments both to the east and to the west.

A secondary possibility is that the anomalous radiogenic lead component started to collect

detrital minerals at the time of their formation ($t_1 = 2500$ m y).

For this assumption we find t_2 to be about 1700 m y. The sulphide minerals have never been considered to have been emplaced recently, and therefore this assumption seems less probable. Alternatively, the anomalous radiogenic lead component could have been produced over a very short interval ($t_2 = t_1$). For this assumption we calculate the time of generation to have been 1450 ± 150 m y ago. An obstacle to this hypothesis is that the production of sufficient radiogenic lead in a very short time would require very large quantities of uranium to have been involved.

Nevertheless this value of 1450 ± 150 m y is important because it is the oldest possible time at which the incorporation of the radiogenic lead could have taken place. That is, if it is correct to place these common lead minerals into a single group as we have done, it is most unlikely that they are as old as the detrital minerals (2500 m y) or even that they are as old as the uraninite (1700 m y), using the term 'age' in the sense of time of last substantial chemical alteration.

CONCLUSIONS

In view of the measurements we have made and considering them in the light of other geological and geophysical evidence, we reach the following conclusions.

Monazite and zircon are known to be detrital minerals. We have dated one of each by the lead-uranium method and have found their ages to be in agreement at 2500 m y. Thus the Lower Mississagi must certainly be less than 2500 m y old. Further evidence has been considered in order to provide a better estimate of its age. The Cutler batholith, which has been considered to intrude the Mississagi has been dated by G. W. Wetherill with discordant results, which however suggest a minimum age in the range of 1300 to 1800 m y. The relationship between the Cutler batholith and the Mississagi has recently been questioned. Ages determined by M. Ozima for biotite and sericite, formed during the metamorphism of the Middle Mississagi by the intrusion of basic rocks, ranged from 1400 to 1600 m y. A granite gneiss within the Bruce which overlies the Mississagi was dated by Ozima at 1720 m y.

Therefore we feel that the age of this sediment probably lies within the range of 1400 to 1800 m y.

The analyses of the uraninite crystals gave ages in agreement at 1700 ± 100 m y. We believe this to be a relatively reliable estimate of age because of the purity of the samples. We believe that the uraninite probably formed within a few hundred years of the time of formation of the sediment.

The age determined for the brannerite is uncertain, but it is probably as old as or older than the uraninite, and significantly younger than the monazite and zircon.

The pyrite, galena, pyrrhotite, and sericite contain lead with isotope ratios which are highly anomalous. The favored interpretation is that the radiogenic lead was derived from uranium minerals formed approximately 1700 m y ago.

The addition of radiogenic lead to supposedly common lead minerals means that this must have been accomplished during recrystallization or introduction of these minerals. This has been shown to have taken place more recently than 1450 ± 150 m y ago, with 1200 to 1300 m y as the best estimate if it is accepted that the lead came from minerals 1700 m y old. In any case, pyrite mineralization was superimposed on the uranium mineralization. The hypothesis of clastic pyrite or simultaneous pyrite and uraninite mineralization becomes valid only if the pyrite has been recrystallized, thereby incorporating the radiogenic lead.

The sericite contains the same type of radiogenic lead as the sulphides, and it presumably formed or recrystallized at the same time. If the sericite could be dated, the time when the radiogenic lead was incorporated into the common lead minerals would be indicated and a means of testing the proposed hypothesis would be provided.

We wish to thank W. Slawson and M. Chinnery for additional chemical and isotopic analyses and to acknowledge the assistance of R. M. Farquhar, F. G. Smith, and W. W. Moorehouse. These studies were made possible by the financial support of the Advisory Committee on Scientific Research, University of Toronto, the National Research Council of Canada, Eldorado Mining and Refining Company of Canada, the Research Council of Ontario, and the Geological

Survey of Canada. The help of Algom Uranium Mines in obtaining samples is acknowledged with thanks.

REFERENCES

- Aldrich, L. T. and G. W. Wetherill, Geochronology by radioactive decay, *Ann. Rev. Nuclear Sci.*, 8, 257-297, 1958.
- Collins, W. H., North shore of Lake Huron, *Geol. Survey Can., Mem.*, 143, 1925.
- Collins, C. B., R. M. Farquhar, and R. D. Russell, The isotopic constitution of radiogenic leads and the measurement of geological time, *Bull. Geol. Soc. Am.*, 65, 1-22, 1954.
- Davidson, C. F., On the occurrence of uranium in ancient conglomerates, *Econ. Geol.* 52, 668-693, 1957.
- Farquhar, R. M., and R. D. Russell, Anomalous leads from the Upper Great Lakes Region of Ontario, *Trans. Am. Geophys. Union*, 38, 552-556, 1957.
- Ginn, R., Ph.D. Thesis, Department of Geological Sciences, University of Toronto, 1959.
- Joubin, F. R., Uranium deposits of the Algoma District, Ontario, *Trans. Can. Inst. Mining M.* 57, 431-437, 1954.
- Maynes, A. D., and W. A. E. McBryde, Determination of traces of lead in igneous minerals, *Anal. Chem.*, 29, 1259, 1957.
- Ozima, M., The dating of minerals by the potassium-argon method, Ph.D. Thesis, Department of Physics, University of Toronto, 1958.
- Roscoe, S. M., Geology and uranium deposits, Quirke Lake, Elliott Lake, Blind River Area, *Geol. Survey Can., Paper* 56-7, 1957.
- Russell, R. D., R. M. Farquhar, G. L. Cummi, and J. T. Wilson, Dating galenas by means of their isotopic constitutions, *Trans. Am. Geophys. Union*, 35, 301-309, 1954.
- Russell, R. D., R. M. Farquhar, and J. E. Hawley, Isotopic analyses of leads from Broken Hill, Australia, *Trans. Am. Geophys. Union*, 38, 545-565, 1957.
- Stanton, R. L., and R. D. Russell, Anomalous leads and the emplacement of lead sulphide ore, *Econ. Geol.*, 54, 558-607, 1959.
- Wanless, R. K., and R. J. Traill, Age of uranium from Blind River, Ontario, *Nature*, 178, 249-249, 1956.

(Manuscript received September 15, 1959.)

Convection Currents in the Earth's Mantle

A. L. LIGHT

*Theoretical Division, Goddard Space Flight Center
National Aeronautics and Space Administration
Washington, D. C.*

Abstract. Recent measurements by J. A. O'Keefe, A. Eckels, and R. K. Squires of the third zonal harmonic component of the earth's gravitational potential are used to determine the characteristics of a hypothetical convection current in the earth's mantle. The calculations are based on a convection theory developed by F. A. Vening Meinesz. The current's velocity is found to be 3.6 cm/yr, corresponding to an overturn time of 1.75×10^8 yr. The temperature perturbation is found to be 9.7°C . The flow is found to impose a distortion of 159 meters on the earth's crust and a distortion of 237 meters on the surface of the core. The mass loadings associated with these distortions were calculated as 5.3×10^7 dyne/cm² at the top, and 8.9×10^7 dyne/cm² at the bottom of the mantle. It is believed that the Vening Meinesz theory is capable of qualitatively explaining the results of O'Keefe and his co-workers. However, the required thermal efficiency is about 70 per cent, as compared with an estimated maximum of 4 per cent. It is considered improbable that convective efficiencies could be so large.

Introduction. The earth's gravitational potential has recently been found by O'Keefe and others [1959] to contain a third-order zonal harmonic component. This component could not be supported by an earth in hydrostatic equilibrium. It is our purpose to give the results of an investigation of the hypothesis that the masses associated with the third harmonic are supported by a system of convection currents in the earth's mantle.

Qualitative discussion. Figure 1 shows in schematic form the sort of convection cell that

might exist in the mantle. The central column of material has been heated, perhaps by conduction from the core. Due to thermal expansion it is lighter than its surroundings and tends to float upwards. The core-mantle interface, normally at the radius R_0 , will be pulled up a height h . The crust-mantle interface, normally at the radius R , will be pushed up a height H . The hot material will overflow at the top, and will cool by conduction to the crust. The surrounding cooler material will encroach upon the hotter column at the bottom, will become heated by conduction and rise; thus the process becomes cyclic.

There is a mass deficiency in the body of the hot column, due to its thermal expansion. There is also a mass excess at the top and bottom of the column, where dense material has been forced into regions of less dense material. The mass excesses will produce positive gravitational anomalies, the deficiencies, negative. The two effects work against each other, and the resultant sign and magnitude of the anomaly will depend critically on the characteristics of the particular flow.

If we suppose that the third-order anomaly discovered by O'Keefe and others [1959] is positive over rising columns (as is true for the particular flow discussed below), then we get a picture like that in Figure 2.

The figure must be thought of as being sym-

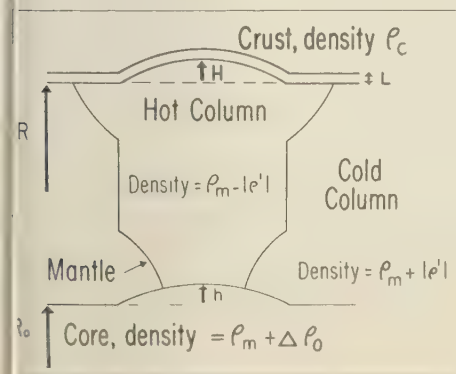


Fig. 1. A schematic representation of a convection cell.

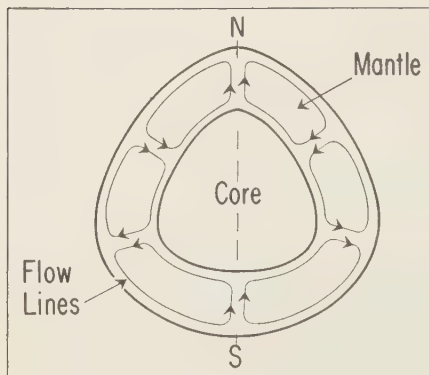


Fig. 2. A third-order convection current

metric about the north-south axis. It is a cross section in a meridional plane of the earth's interior, if a third-order convection current is assumed to be present. Both the density perturbation and the radial component of the convecting velocity vary as the third-order zonal harmonic. The current rises under the North Pole, falls at approximate latitude 25°N , rises again at latitude 25°S , and finally descends under the South Pole.

The gravitational perturbation. For such a flow, the perturbation in the gravitational potential $U'(\mathbf{r})$, considered at the point \mathbf{r} of radius r and direction $\mathbf{n} = \mathbf{r}/r$, is easily obtained from the well-known expression

$$U'(\mathbf{r}) = G \int \frac{\rho'(\mathbf{r}') d\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} \quad (1)$$

If we expand U' in terms of the spherical harmonics $Y_{kl}(\mathbf{n})$,

$$U'(\mathbf{r}) = \sum \frac{A_{kl} Y_{kl}(\mathbf{n})}{r^{k+1}} \quad (2)$$

then the coefficient A_{kl} can be shown to satisfy, to the first order in small quantities,

$$A_{kl} = \frac{4\pi G}{2k+1} \left(R^{k+2} \rho_m H_{kl} + R_0^{k+2} \Delta \rho_0 h_{kl} + \int_{R_0}^R r'^{k+2} \rho_{kl}'(r') dr' \right) \quad (3)$$

Essentially, this is just a k 'th moment of the mass distribution. G is the universal constant of gravitation, and the rest of the quantities are

defined as in Figure 1. The subscripts indicate that a spherical harmonic expansion is to be made of the appropriate quantity, and the coefficient used in (3).

There are three terms that contribute to A_{kl} . The first is due to the heaping up at the top of the mantle, the second to the pulling up of the core, and the third to the density perturbation within the mantle.

From the work of *O'Keefe and others* [1956] we know that

$$A_{30} = 0.24(\text{megameters})^6/(\text{kiloseconds})^2 \quad (4)$$

This provides an external datum to which we must fit any convection theory which professes to give a current distributed according to the third-degree zonal harmonic. If the theory were to provide the relative magnitudes of the quantities entering into (3), then we could use (4) to determine their absolute magnitudes.

The Vening Meinesz convection theory. It has been found possible to fit to (4) a convection current predicted by a theory developed by F. Z. Vening Meinesz [*Heiskanen and Vening Meinesz*, 1958; this reference will be referred to hereafter as HVM].

In this theory the mantle is treated as a classical viscous fluid. Possible changes in phase and chemical composition are neglected. It is doubtful that such an idealization could represent the real mantle in any but a very approximate way. Unfortunately, a more realistic development of the convection hypothesis is not at present available.

Calculations. On page 442 of HVM, expressions are given for the vertical velocity w in dimensionless units as a function of $\epsilon = r/l$ for each harmonic of order k . That is, we have

$$w_{kl} = D_{kl} K_k(\epsilon) \quad (5)$$

where D is a constant with the dimensions of velocity, which we shall evaluate with the aid of (3), and K_k are the functions formulated in HVM, page 442. We have calculated these functions and plotted them in Figure 3.

Now the quantities $\rho_m H$, $\Delta \rho_0 h$ that enter into (3) are simply the mass surface densities of the material heaped up at the top and bottom of the mantle. Such a mass load may be expressed in terms of the derivatives of w by equating, with appropriate choice of sign, its weight to

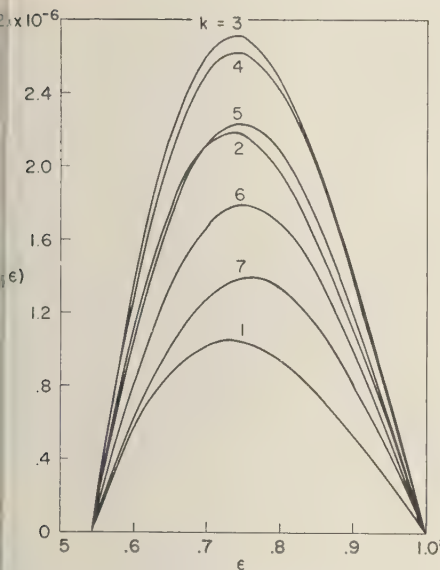


Fig. 3. Vertical velocity amplitudes $K_k(\epsilon)$, as functions of radius ϵ for harmonics of order $k = 1$ to 7. K_k is in arbitrary units, ϵ is in units of the Earth's radius.

the rr component of the stress tensor associated with the motion at the appropriate surface.

We find from HVM (equations 28 and 43) and equation (5) that

$$H_{kl} = -\frac{\eta}{gR} T_k^1; \quad \Delta \rho_0 h_{kl} = +\frac{\eta}{gR} T_k^0 \quad (6)$$

where the stress, in dimensionless units, is

$$\left. \begin{aligned} & \frac{1}{\epsilon^2} K_k'''(\epsilon) \right\} \begin{aligned} & \text{at } \epsilon = 1 \\ & \text{at } \epsilon = 6/11 \end{aligned} \\ & - \epsilon^2 K_k'''(\epsilon) \end{aligned}$$

where η is the mantle's viscosity, g is the acceleration of gravity, and primes denote $d/d\epsilon$.

The integral over the density perturbation that appears in (3) can be expressed as

$$-\alpha \rho_m \int_{R_0}^R r^{k+2} \Theta_{kl}(r') dr'$$

where α is the coefficient of thermal expansion, and Θ is the temperature perturbation associated with the convection. This is given as

$$\Theta_{kl} = C_{kl} F_k(\epsilon) \quad (7)$$

The F_k are the functions formulated in HVM, page 416. C_{kl} is defined in terms of D_{kl} by HVM (equation 37) as

$$C_{kl} = \frac{\eta}{\alpha \rho_m g R^2} D_{kl} \quad (8)$$

Figure 4 is a plot of the F_k , for k from 1 to 7. Thus we have

$$\int r^{k+2} \rho_{kl}'(r') dr' = -\frac{\eta}{g} R^{k+1} D_{kl} E_k \quad (9)$$

where

$$E_k = \int_{6/11}^1 \epsilon^{k+2} F_k(\epsilon) d\epsilon$$

Equation (3) can now be reduced to

$$A_{kl} = \frac{4\eta G}{2k+1} \frac{\eta}{g} R^{k+1} \left(T_k^1 - \left(\frac{6}{11} \right)^{k+2} T_k^0 - E_k \right) D_{kl} \quad (10)$$

The numbers T_3^1 , T_3^0 , E_3 have been evaluated numerically.

$$T_3^1 = +0.781 \times 10^{-4}.$$

$$T_3^0 = -1.321 \times 10^{-4}.$$

$$E_3 = +0.566 \times 10^{-4}. \quad (11)$$

Using (11), the value (4) for A_{30} , and Vening Meinesz's value of 10^{22} poise for η , we obtain

$$D_{30} = 0.0428 \text{ cm/sec} \quad (12)$$

From Figure 3 we see that

$$(w_{30})_{\max} = 2.70 \times 10^{-6} D_{30}$$

by (12), this is $(w_{30})_{\max} = 1.15 \times 10^{-7}$ cm/sec = 3.64 cm/yr. We can associate with this velocity a time of overturn T :

$$T = R/w = 175 \times 10^6 \text{ yr}$$

From Figure 4,

$$(\Theta_{30})_{\max} = 0.61 \times 10^{-3} \frac{\eta}{\alpha \rho_m g R^2} D_{30}$$

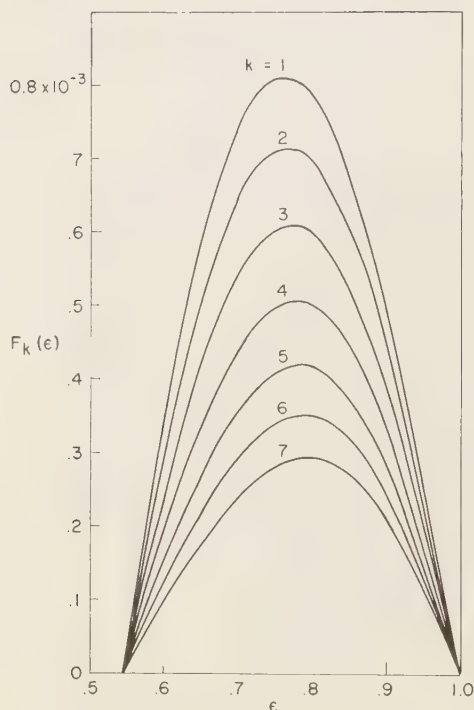


Fig. 4. Temperature perturbation amplitudes $F_k(\epsilon)$, as functions of radius ϵ , for harmonics of order $k = 1$ to 7. Units are as in Figure 3.

With the standard values

$$\rho_m = 3.3 \text{ g/cm}^3 \quad \text{and} \quad \alpha = 2 \times 10^{-5} \text{ } ^\circ\text{C}^{-1}$$

we find,

$$(\Theta_{30})_{\max} = 9.7 \text{ } ^\circ\text{C}$$

The weights of the heaped-up material are, from (6),

$$g\rho_m H_{30} = 5.3 \times 10^7 \text{ dyne/cm}^2;$$

$$g\Delta\rho_0 h_{30} = 8.9 \times 10^7 \text{ dyne/cm}^2$$

from which we find, taking $\Delta\rho_0 = 3.75 \text{ g/cm}^3$, that

$$H_{30} = 159 \text{ m}$$

$$h_{30} = 237 \text{ m}$$

Discussion. The magnitudes of these results depend critically on the value chosen for the viscosity η . Vening Meinesz's figure of 10^{22} poise

was determined from the rising of the Fennoscandia basin. *Lyustikh* [1957], however, claims that a similar calculation based on the sinking of the Podmoskonmaia basin gives $\eta = 10^{26}$ poise. As may be seen from equations (10) and (5), this would reduce the velocity by a factor of 10^4 and the time of overturn would be increased to 1.75×10^{13} year. Such a movement would be insignificant. The displacement since the creation of the earth would amount to, at most, 18 km.

The 3.6 cm/yr velocity is of roughly the same order of magnitude as that calculated by previous authors, using about the same viscosities. *Scheidegger* [1952] derived a value of 2 mm/yr from *Bullard's* [1950] estimate of the outflow heat from the core. *Pekeris* [1935] derived, from a variety of assumptions, velocities ranging from 0.1 to 20 cm/yr.

It is possible to calculate the efficiency of the flow. The power P lost through viscous dissipation is roughly

$$P \approx \eta(w/L)^2$$

If we take L to be one-half the depth of the mantle, $1.45 \times 10^8 \text{ cm}$, then

$$P \approx 0.63 \times 10^{-8} \text{ erg/cm}^2 \text{ sec}$$

Multiplying this by the volume of the mantle, $9.2 \times 10^{26} \text{ cm}^3$ we get an estimate of the rate at which the flow does work against the viscous forces:

$$R \approx 5.8 \times 10^{18} \text{ erg/sec}$$

A value for the efficiency may be obtained by comparing this with the power available through conduction from the core. *Jacobs* [1956] gives two estimates, 3.6 and 4.2 cal/cm² yr, for the heat flow from the core, based on two different models. We will take as a mean 4.0 cal/cm² (5.3 erg/cm² sec). Multiplying this by the surface area of the core, $1.54 \times 10^{18} \text{ cm}^2$, we get an estimate of the power available, P_a ,

$$P_a \approx 8.2 \times 10^{18} \text{ erg/sec}$$

The ratio R to P_a , the efficiency required for the convecting system, is about 70 per cent. We wish to compare this with an estimate of the maximum possible efficiency.

In *Jacob's* model I [1956], the surface of the core is at a temperature of 4282°C, and the ba-

the crust is at 461°C. The maximum efficiency of an ideal heat engine running between these two temperatures is 84 per cent. To obtain an estimate of the maximum efficiency of a real heat engine, we must multiply this by the appropriate coefficient of utility, n . For a convective system a reasonable estimate for n is

$$n = (B - B_0)/B$$

where B is the actual and B_0 the adiabatic thermal gradient. Using a value of 6×10^{-3} °C/cm sec for the thermal conductivity, we can estimate B from the above rate of heat flow as

$$B = 0.21 \times 10^{-4} \text{ °C/cm}$$

A calculation based on the Vening Meinesz theory gives

$$B - B_0 = 0.61 \times 10^{-10} \text{ °C/cm}$$

and therefore $n = 3 \times 10^{-6}$. This yields a maximum efficiency of 2.4×10^{-4} per cent. Hales (1935), using a somewhat different approach, found values of $B - B_0$ ranging from 10^{-6} to 10^{-9} °C/cm, for viscosities ranging from 10^{25} to 10^{28} . Even if we use Hales' figures, the efficiency is not more than 4 per cent. These values are much smaller than the 70 per cent efficiency required by the Vening Meinesz theory of convection.

Conclusion. We conclude that the Vening Meinesz theory of convection can account qualitatively for the results of O'Keefe and others

[1959]. However, if we accept the Vening Meinesz value for the viscosity of the mantle, the required thermal efficiency becomes too high for a convective process. It remains to be seen whether a more realistic theory of convection with a lower efficiency could account for the observed A_{10} .

The author wishes to express his thanks to J. O'Keefe and R. Jastrow for many helpful discussions and to G. MacDonald for his stimulating criticisms.

REFERENCES

- Bullard, E. C., The transfer of heat from the core of the earth, *Monthly Notices Roy. Astron. Soc., Geophys. Suppl.*, 6, 36-41, 1950.
- Hales, A. L., Convection currents in the earth, *Monthly Notices Roy. Astron. Soc., Geophys. Suppl.*, 3, 372-379, 1935.
- Heiskanen, W. A., and F. A. Vening Meinesz, *The Earth and Its Gravity Field*, chap. 11, McGraw-Hill, New York, 1958.
- Jacobs, J. A., The earth's interior, pp. 398, in *Handbuch der Physik*, vol. XLVII, Springer-Verlag, Berlin, 1956.
- Lyustikh, E. N., *Izvest. Akad. Nauk SSSR, Ser. Geofiz.*, 5, 604-615, 1957.
- O'Keefe, J. A., A. Eckels, and R. K. Squires, *Science*, Feb., 1959.
- Pekeris, C. L., Thermal convection in the interior of the earth, *Monthly Notices Roy. Astron. Soc., Geophys. Suppl.*, 3, 343-367, 1935.
- Scheidegger, A. E., Physical aspects of the convection-current hypothesis of orogenesis, *Trans. Am. Geophys. Union*, 33, 585-586, 1952.

(Manuscript received October 4, 1959.)

Letters to the Editor

Soil Moisture Recounting Under a Permanent Grass Cover¹

M. A. HARTMAN

*Blacklands Experimental Watershed
Riesel (Waco), Texas*

This is a discussion of one method and pertinent factors used in computing soil moisture from climatological data. The basic concept, factors, and methods involved were very aptly discussed by *Smith* [1959].

In general, it is recognized that methods for estimating soil moisture from readily available information have deficiencies [*van Bavel*, 1956]. Percolation below the zone under consideration (18 inches, *Smith* [1959]) and surface runoff may affect the relationships significantly. It is assumed that these two items are not important factors at Imperial College in Trinidad because *Smith's* [1959] statement that 'the difference between this amount (evapotranspiration) and

the rainfall gives the quantity of water remaining in the soil.'

The importance of deep percolation and surface runoff, at least for some conditions, in determination of soil moisture by a water balance accounting is illustrated in Table 1. These data are from a 3-acre Bermuda grass (*Cynodon dactylon*) pasture, grazed intermittently and in fair vegetative condition. The soils of this small watershed are representative of the soils of the Texas Blacklands area, and are mostly deep, heavy montmorillonite clay soils which crack severely when dry. The evaporation from the Young screened evaporation pan is indicative of evaporation from the water surface [*Bloodgood and others*, 1954].

A brief review of these data indicates that accretions and depletions from the top 3 feet are not indicative of the changes in the fourth and

¹ Contribution from Watershed Technology Research Branch, Soil and Water Conservation Research Division, Agricultural Research Service, U. S. Department of Agriculture.

TABLE 1. Soil Moisture Relationships in a Bermuda Grass Pasture

Date	Soil Moisture,* inches			Evapo- ration, inches †	Precipi- tation, inches ‡	Runoff, inches §
	Top foot	Top 3 ft.	Top 5 ft.			
1957						
April 12	1.63	4.99	6.61
May 6	2.66	6.81	9.02	2.02	15.22	10.62
May 22	1.67	6.09	9.90	1.69	6.60	4.64
1958						
Aug. 5	0.06	1.57	4.31
Aug. 26	1.89	4.70	7.07	5.73	6.33	0.06
Sept. 18	1.27	3.61	5.64	4.08	1.73	0
Oct. 16	1.39	4.65	7.66	3.91	4.01	0.24

* Soil moisture in inches above 18 per cent of dry weight (approximate wilting point). Determined by gravimetric methods from sampling by increments with auger at three locations. (Apparent specific gravities: 1.2 to 6 inches, 1.2; 6 to 12 inches, 1.3; 12 to 36 inches, 1.4; 36 to 60 inches, 1.5.)

† Accumulated evaporation between dates from daily readings of losses in a 24-inch Young screened evaporation pan.

‡ Accumulated precipitation between dates, from standard rain gage.

§ Accumulated runoff between dates, from H-3 measuring flume with stage recorder.

fifth feet. For example, the top 3 feet gained about 1 inch from September 18 to October 16, 1958, whereas the top 5 feet gained about 2 inches. This means a gain of 1 inch in the 3- to 5-foot layer for this period. The period from August 5 to August 26, 1958, shows a gain of about 3 inches for the top 3 feet and a loss of 0.37 inch for the fourth and fifth feet. The effect that surface runoff may have on soil moisture determinations is obvious for the 1957 periods. During this 40-day period, 70 per cent of the available supply in the form of rainfall was surface runoff and therefore not available for soil moisture storage on the area under consideration.

In making estimates of flood runoff from ungaged watersheds the amount of moisture in the soil is a most significant factor, for it is an indicator of the amount of rainfall that can be retained on a watershed. In order to provide a soil-moisture index for use in determining runoff, a method of estimating soil moisture has been developed by the writer and his associates. The estimate of soil moisture is computed by an accounting system using land uses, daily rainfall, and pan evaporation data [Hartman and others, 1959]. The pan evaporation data represent an integration of many of the forces involved in the evaporative processes.

The change in soil-moisture dissipation, including evaporation, transpiration, and deep percolation, is expressed as a decay-type curve, that is, a curve decreasing in a manner such that the incremental change is directly proportional to the amount of soil moisture [Slayter, 1956]. This type of curve can be expressed by the equation

$$SM_t = SM_0 K^t \quad (1)$$

where

SM_t = soil moisture t days after the last observation

t = number of days since last observation

K = the dissipation constant or ratio of the soil moisture on any day to the previous day's soil moisture

Using the parameters of rates of soil-moisture dissipation, amount of soil moisture, and pan evaporation, seasonal relationships by land uses can be developed by multiple regression methods.

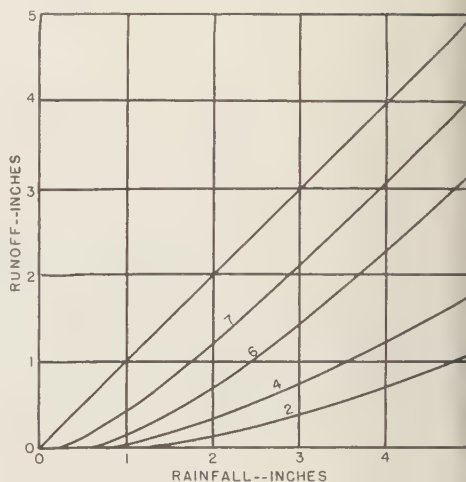


Fig. 1. Daily Rainfall versus Runoff with various ASM indices for Bermuda grass pasture near Riesel (Waco), Texas.

The equations of these relationships for Bermuda grass pasture are

$$K_s = 0.958 + 0.003SM_0 - 0.012PE \quad (2)$$

$$K_w = 0.990 + 0.001SM_0 - 0.033PE \quad (3)$$

where

K_s = dissipation constant for March to October 15

K_w = dissipation constant for October 15 to March

SM_0 = soil moisture above 18 per cent in 5 feet at end of last day's observation (inches)

PE = average daily evaporation from the Young screened pan for the period (inches)

In the runoff studies carried out by the writer runoff is treated as a residual of rainfall after abstractions; that is, the amount of daily runoff is the difference between daily rainfall and the daily amount of abstractions or retentions. There is a significant correlation between the amount of moisture in the soil before the rain and the runoff. The essential information is shown in Figure 1, which for Bermuda grass pasture on Houston black clay represents the

relationship between daily rainfall and runoff, with antecedent soil moisture, ASM, as a retention factor.² The ASM index is equal to the inches of water greater than 18 per cent by dry weight in the top 3 feet of soil. With the use of this figure and the soil moisture before the rain begins, determined by equation (1), runoff from any given amount of daily rainfall can be estimated. The rainfall minus computed runoff is added to the computed initial soil moisture to obtain the moisture in the soil after the rain.

This method has inherent characteristics that reduce and compensate for the effect of errors. An overestimate of soil moisture will result in an underestimate of water added; thereby the error of the estimate of soil moisture is reduced.

² Forthcoming *Texas Agr. Expt. Sta. Bull.* by Hartman and others.

REFERENCES

- Van Bavel, C. H. M., Estimating soil moisture conditions and time for irrigation with the evapotranspiration method, *U. S. Dept. Agr., ARS 41-11*, 16 pp, August 1956.
- Bloodgood, Dean W., R. E. Patterson, and R. L. Smith, Jr., Water evaporation studies in Texas, *Texas Agr. Expt. Sta. Bull.* 787, 83 pp, November 1954.
- Hartman, M. A., R. W. Baird, J. B. Pope, and W. G. Knisel, Jr., Prediction of soil moisture from rainfall and pan evaporation records, *Southern Agricultural Workers' Association, February 1959*, Memphis, Tennessee.
- Slayter, R. O., Evaporation in relation to soil moisture, *Neth. J. Agr. Sci.*, 1, 73-76, 1956.
- Smith, G. W., The determination of soil moisture under a permanent grass cover, *J. Geophys. Research*, 64, 477-483, 1959.

(Received July 22, 1959; received October 10, 1959.)

Criticism of the Melted Moon Theory

HAROLD C. UREY

*University of California
La Jolla, California*

Kuiper [1954] has proposed that the moon accumulated at low temperature and was heated to the melting point by the decay of radioactive elements, until the interior melted, leaving a sintered 'accreted crust' on the outside. At this stage massive objects collided with the moon and broke through the crust, and lava from the interior flowed up into the holes thus formed. On solidification, these pools of lava formed the maria and the flooded craters (referred to as the 'wet craters'). Later the moon solidified, and subsequent collisions produced the 'dry craters,' i.e., those not having flooded interiors.

A basic objection may be raised against this hypothesis. *Kuiper* [1959] in a recent publication estimates the thickness of this accreted crust as 16 km, below which it is assumed that there is an unspecified but presumably substantial layer of liquid. He apparently assumes that the accreted matter was uncompacted finely divided material. The pressure at 16 km in the moon, assuming a density of the surface materials of 2.5 g cm^{-3} , is 660 atmospheres, sufficient to compact the material at this depth, and the pressures at shallower depths would do so also. For any reasonable assumption on the composition of the moon the compacted solid will be denser than the liquid. Solid silicate, for example, is denser than melted silicate by 4 to 12 per cent. In this circumstance, in which all but the surface of the layer of the crust can be expected to have a higher density than the liquid, the configuration is unstable, and in any disturbance the crust must break up and sink into the liquid.

In this connection we quote *Kelvin* [1876]: 'Suppose the earth this moment to be a thin crust of rock or metal resting on liquid matter; its equilibrium would be unstable! And what of the upheavals and subsidences? They would be strikingly analogous to those of a ship which has been rammed—one portion of crust up

and another down, and then all down. Now whatever may be the relative densities of rock solid and melted, at, or about, the temperature of liquefaction, it is, I think, quite certain that cold solid rock is denser than hot melted rock; and no possible degree of rigidity in the crust could prevent it from breaking to pieces and sinking wholly below the liquid lava.'

It is possible that a liquid core might have a higher density than the solid if we allow for the fact that iron oxide concentrates in the liquid phase relative to the solid phase in a solid-liquid equilibrium involving a silicate. Using the Bowen-Schairer data I have estimated that the solid will float on a liquid core if the core contains some 35 per cent of iron oxide. However, this iron content or even a smaller one would be inconsistent with the moon's observed density unless several per cent of water were present in the moon. But water would concentrate in the melt and lower its density, and hence no floating of the solid could occur. Also the presence of other elements, the alkalis and aluminum and others, would change the calculated results and would act in the direction of lower densities for the liquid phase. Even extreme assumptions cannot give a lower density for the solid phase.

It is an observed fact that the solid crystals always sink during solidification of silicate melts containing high concentrations of iron oxide e.g., as in the Skergaard intrusion.

These calculations, however, do not apply to *Kuiper's* model, for he assumes that an original 'accreted crust,' which would have average primitive composition, floats on a completely melted moon below this crust, which therefore also has an average primitive composition. In this case the solid is more dense than the liquid and the crust must disintegrate and sink into the liquid.

There are several other difficulties in the melted moon hypothesis:

1). Calculations on the heat to be expected from radioactive decays indicate that an age greater than 5.5 billion years must be assumed in order to produce the melting according to the postulates of Kuiper, as nearly as this reader can judge [Urey, 1956]. But present estimated values for the ratio of lead to uranium require that the age cannot be greater than 5.5 billion years.

2). The irregular shape of the moon implies a substantial mechanical strength, which would seem to be incompatible with the melted moon model. If, as Kuiper suggests, the irregularity developed through the fall of objects on the moon at a time when the interior of the moon was completely melted, the uncompensated differences again could not be maintained over several subsequent eons. The melted moon hypothesis also is difficult to reconcile with arguments first put forward by Gilbert showing that the outer parts of the moon must have been relatively cold and very rigid at the time the maria were formed, in order that the very substantial differences in elevation could be maintained. Watts reports a difference in elevation on the east limb of 9.7 km. On the earth such uncompensated differences in height could not remain for even 10^6 years.

3). The melting process assumed for the moon is unrealistic in the light of our knowledge of corresponding terrestrial processes. If the moon is accreted at low temperature it should partly melt by radioactive heating, producing a liquid of low density containing greatly

increased concentrations of potassium, uranium, and thorium. This liquid should flow toward the surface at some stage, and either flow out on the surface or under a loosely compacted surface of low density. In either event the liquid would lose heat to the environment by conduction and radiation and solidify quickly. Other flows should follow until the radioactive elements were removed from the deep interior. This is what appears to have happened to some extent on the earth.

Kuiper's model is inconsistent with discussions published in the scientific literature by scientists many years ago [Gilbert, 1893; Urey, 1951; and others]. What is more important, the model is inconsistent with very elementary physical and chemical principles, and well-established observations.

REFERENCES

- Gilbert, G. K., *Bull. Phil. Soc. Wash.*, 12, 241-292, 1893.
- Kelvin, *Rept. Brit. Assoc. Advance. Sci.*, pp. 1-12, 1876.
- Kuiper, G. P., *Proc. Natl. Acad. Sci. U. S.*, 40, 1096, 1954.
- Kuiper, G. P., *Vistas in Astronautics*, vol. II, pp. 273-312, 1959.
- Urey, H. C., *Geochim. et Cosmochim. Acta*, 1, 207, 1951.
- Urey, H. C., *Proc. Natl. Acad. Sci. U. S.*, 42, 889-891, 1956.
- Urey, H. C., W. Elsasser, and M. G. Rochester, *Astrophys. J.*, May 1959, and Meetings of Natl. Acad. Sci., Washington, April 1959.

(Received November 5, 1959.)

Polar Auroral, Geomagnetic, and Ionospheric Disturbances

E. H. VESTINE

*The Rand Corporation
Santa Monica, California*

In a recent letter Vestine and Sibley [1959] noted that the integral invariant of auroral particle motion

$$I = \int_{mN}^{ms} \sqrt{1 - F/F_m} dl \quad (1)$$

appears to predict successfully features associated with auroral isochasms, when the geomagnetic field in space can be computed. In equation 1, dl is an element of length of a geomagnetic line of force along which the charged particle is spiraling while slowly drifting to adjacent lines. In the integral, F is the geomagnetic field strength, and F_m the field value at which the particle is reflected (the mirror point). In fact $(1 - F/F_m)^{1/2} v = v_p$ is regarded as the approximate component of the velocity parallel to the field line, where v is the total velocity of the particle. This integral was derived for charged-particle motion in a magnetic field by Rosenbluth and Longmire [1957], and later discussed by others.

It is the purpose here to suggest that this simple concept may be useful in explaining a number of other polar effects that have remained unexplained for a long time. These include the increased frequency and intensity of aurora near midnight, as well as the two polar electrojets associated with polar magnetic disturbances and radio blackouts, and the tendency of magnetic bays to recur near the same universal hour of time for several consecutive nights.

In equation 1 it will be noted that the quantity $(1 - F/F_m)^{1/2}$ can never exceed unity, and that, since F tends to diminish with distance from the earth, both this function and the length of line should increase if a geomagnetic field line is stretched in space, owing to the hydromagnetic action of solar streams discussed later in this note. Hence the value of I for a particular field line may increase with stretching of the field line and diminish with the compression of the

field which may shorten a field line. If, then, a particular line of force is compressed or stretched, the value of I can be expected to change.

If a solar stream of protons is imagined, or clouds of protons impinging on one side of the earth, the geomagnetic field may be compressed, thereby increasing F , and the field lines may possibly be shortened, on the sunward side of the earth. As is well known, if the kinetic energy of the stream F is $1/2nmv^2$, where n is the number density of particles, m the mass of each particle, and v the velocity, and if this kinetic energy is greater than the magnetic energy $F^2/8\pi$ of the local field, the stream will carry the lines of force along or stretch them to greater distances. If the stream proceeds directly from the sun, some compression of the outer geomagnetic field on the noon meridian will result, and stretching of lines of force will occur on the night side. In fact, an illustrative case for the initial condition has been calculated by Chapman and Ferraro [Chapman and Bartels, 1940] in their theory of the initial phase of storms. As the stream advances, drainage of particles into the geomagnetic field on the sunlit side has often been suggested, either as clouds of particles, beams, as suggested by Bennett [1959], or otherwise. Those impinging on the day side which might penetrate the field should spiral down to their mirror points along field lines compressed by the solar stream. There will be a value of I appropriate to a particular particle mirroring on the surface F_m . The height of this surface will be greatest in polar regions and will decrease equatorward.

As the particle drifts slowly to the west (if it is a proton), it must do so in accordance with equation 1. This means that, if the field lines near the dawn meridian are stretched to a length greater than that near noon by parts of the stream at higher levels passing the earth, the particle probably must move equatorward to a

of shorter length in order that equation 1 be satisfied. If the shift equatorward in the reflection point is great enough, the mirror point r_m may be so reduced in height that it is at levels of the polar atmosphere dense enough to provide aurora. On this basis the auroral particles should penetrate more deeply into the atmosphere at night. It may be that the flux density will be greater near the westward-directed auroral electrojet associated with magnetic bays. In fact, *Heppner* [1954] and many others have noted the close association of strong magnetic bays occurring near midnight and early morning hours with overhead aurora. The drift of auroral protons (or electrons) in accordance with curves of equal I [*Vestine and Sibley*, 1959] would be in the right sense.

Application of the integral invariant to problems of the kind discussed here should be regarded as well founded and sound. It may, however, be appropriate to note another physical factor, though in the absence of detailed calculations its importance is difficult to assess. It may be that the solar streams are of large cross section, sweeping across the earth over a period of, say, 3 or 4 days as the sun rotates on its axis. In this way the tendency of polar magnetic bays to recur nightly in a 3- or 4-day sequence might be explained as a unitary phenomenon, penetration of particles to the polar regions occurring under the special conditions discussed above.

In other words, the mirror height for a group of particles of given pitch angle and energy is localized and might therefore explain the tendency of magnetic bays to recur (most intensely near the auroral zone) near the same locality at the same hour for several consecutive nights [*Chapman and Bartels*, 1940].

Such a stream might have particles moving past the earth along paths parallel to stretched field lines far above the auroral zones. It would seem that, since mobility is very high parallel to the lines, the trapping of particles might be facilitated, through drift in the direction $\mathbf{F} \times \text{grad } F$ in the distorted field. These particles would then drift to lines with a similar value of I . If on the dawn or night side the stretching of lines by the stream is a little greater, the particles may mirror farther south (in the northern hemisphere), penetrating the lower ionosphere

at the auroral zone. In the same way, trapping may also be facilitated on the evening side, the particles drifting to field lines of value I so that penetration of the atmosphere occurs. In this way two polar electrojets at the auroral zone may arise, one in the evening and one on the early morning side, as required by observation, though it is uncertain whether the electric field generated by particles absorbed by the atmosphere on the evening side will be in the right sense.

Another possibility is that the eastward-directed electrojet at the auroral zone might arise on the afternoon and evening side of the earth from the portion of directly trapped solar gases (protons and electrons) with smaller particle pitch angle and low mirror point. On the other hand, the westward-directed electrojet might arise from the trapped particles of larger pitch angle in a region where a low mirror point results, in accord with equation 1, because of stretched field lines. Calculations of resulting charge separations low in the ionosphere might reveal whether the electric driving forces (though now correctly directed) are adequate locally near each jet. In this case the initial trapping of the particles would occur on the afternoon or evening side of the earth, perhaps because the outermost lines of force of the geomagnetic field would be carried along by the solar stream, so that auroral regions are openly accessible to the following parts of the stream.

Conversely, since auroral displays and geomagnetic bays are known to tend to occur more or less simultaneously in the northern and southern hemispheres, it may be inferred that precipitation of trapped particles occurred in accord with equation 1. The trapping function of the geomagnetic field need only be temporary, and may terminate because of absorption of particles in the low ionosphere, or by the incapacity of the geomagnetic field to contain them.

This work was done with the aid of an IGY grant of the National Science Foundation.

REFERENCES

- Bennett, Willard H., Solar proton stream forms with a laboratory model, *Rev. Sci. Instr.*, 30, 64-69, 1959.

Chapman, S., and J. Bartels, *Geomagnetism*, Oxford University Press, 1940.

Heppner, J. P., Time sequences and spatial relations in auroral activity during magnetic bays at College, Alaska, *J. Geophys. Research*, 69, 329-338, 1954.

Rosenbluth, M. N., and C. L. Longmire, Stability

of plasmas confined by magnetic fields, *Ann. Phys.*, 1, 120-140, 1957.

Vestine, E. H., and W. L. Sibley, Remarks on auroral isochasms, *J. Geophys. Research*, 64, 1338-1339, 1959.

(Received November 2, 1959.)

A Note on the Effect of Surface Loading on the Shear Response of Overburdens

NICHOLAS N. AMBRASEYS

*Department of Civil Engineering
Imperial College of Science
London S. W. 7, England*

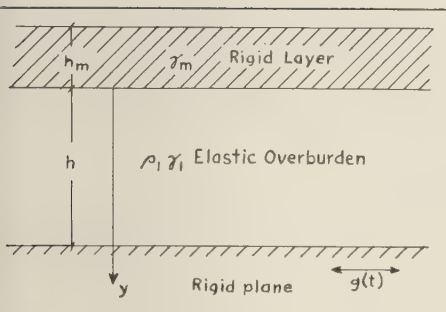


Fig. 1. Diagrammatic representation of the load-overburden system.

In this note we shall give briefly the solution to a problem which often arises in applied seismology. The problem can be stated as follows: to find the shear response characteristics of an elastic homogeneous overburden subject to a seismic base disturbance when its surface is uniformly loaded.

The solution of this problem should furnish

the free periods of coupled oscillations for the overburden-load system, shear stresses in the overburden, and of the magnification of the bedrock motion.

Let it be assumed that an elastic layer of thickness h and density γ is in non-slip contact, at its base with a rigid plane and at its surface with a rigid layer of thickness h_m and density γ_m (Fig. 1). Let it also be assumed that from $t = 0$ the rigid plane executes horizontal oscillations, the amplitudes of which are given by an arbitrary function of time, $g(t)$. Referring to the system of coordinates shown in Figure 1, we see that the governing equation of the motion is given by

$$\ddot{u}(y, t) + c\dot{u}(y, t) - g_0(t) = s^2 \frac{\partial^2 u(y, t)}{\partial y^2} \quad (1)$$

in which, $u(y, t)$ denotes the relative displacement of any point in the medium, c the damping constant of the material, s the velocity of shear waves, and $g_0(t) = -\ddot{g}(t)$.

The boundary conditions may be introduced

TABLE 1. Roots of the transcendental equation $\mu a_r \tan(a_r) = 1$

μ	a_1	a_2	a_3	a_4	a_5	a_6
0	1.5708	4.7124	7.8540	10.9956	14.1372	17.2788
0.1	1.4289	4.3058	7.2281	10.2003	13.2142	16.2594
0.2	1.3138	4.0336	6.9096	9.8928	12.9352	16.0107
0.5	1.0769	3.6436	6.5783	9.6296	12.7223	15.8336
1.0	0.8603	3.4256	6.4373	9.5293	12.6453	15.7713
2.0	0.6533	3.2923	6.3616	9.4775	12.6060	15.7397
5.0	0.4328	3.2039	6.3148	9.4459	12.5823	15.7207
10.0	0.3111	3.1731	6.2991	9.4354	12.5743	15.7143
25.0	0.1987	3.1543	6.2895	9.4290	12.5696	15.7105
50.0	0.1410	3.1479	6.2864	9.4269	12.5680	15.7092
100.0	0.0998	3.1448	6.2848	9.4258	12.5672	15.7086
1000.0	0.0316	3.1419	6.2833	9.4249	12.5665	15.7080
∞	0	3.1416	6.2832	9.4248	12.5664	15.7080

TABLE 2. Values of $D(\mu, r)$

μ	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$
0	+1.275	-0.425	+0.254	-0.182	+0.142	-0.116
0.1	1.260	0.393	0.213	0.130	0.089	0.063
0.2	1.240	0.345	0.158	0.088	0.0545	0.0364
0.5	1.180	0.238	0.085	0.0413	0.0242	0.0156
1.0	1.120	0.153	0.047	0.0216	0.0123	0.0080
2.0	1.070	0.088	0.0244	0.0110	0.0063	0.0040

n the form

$$u(h, t) = 0, \quad t > 0$$

$$u(y, 0) = 0$$

$$\dot{u}(y, t) = 0, \quad t = 0$$

and

$$\rho s^2 \frac{\partial u(y, t)}{\partial y} = \frac{\gamma_m h_m}{g} [\ddot{u}(y, t)$$

$$+ c\dot{u}(y, t) - g_0(t)], \quad y = 0$$

where the quantity in the brackets is the absolute acceleration of any particle in the medium.

Denoting by λ the fraction of critical damping, by μ the quantity $(\gamma_m h_m / \gamma h)$, and by a_r the r th root of the transcendental equation

$$\mu a_r \tan(a_r) = 1 \quad (3)$$

it can easily be shown that the solution of (1) which satisfies the boundary conditions (2) is

$$u(y, t) = - \sum_{r=1,2,\dots}^{\infty} D(\mu, r) D(y, \mu, r) S_d \quad (4)$$

with

$$D(\mu, r) = 2[\mu a_r^2 \cos(a_r) + a_r(1 + \mu) \sin(a_r)]^{-1}$$

$$D(y, \mu, r) = [\cos(a_r y/h) - \mu a_r \sin(a_r y/h)]$$

$$S_d = \omega_r^{-1} (1 - \lambda^2)^{-1/2}$$

$$\cdot \int_0^t g_0(T) \exp[-\omega_r \lambda(t - T)]$$

$$\cdot \sin \omega_r (1 - \lambda^2)^{1/2} (t - T) dT$$

where ω_r is the undamped natural frequency of the system which is given by

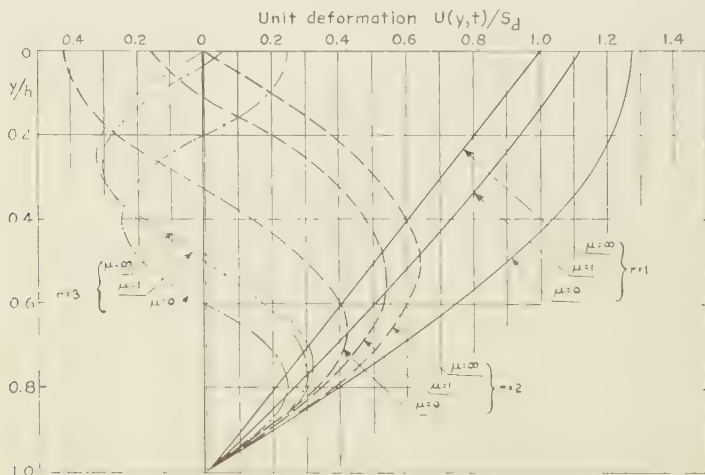


Fig. 2. Effect of load factor μ on unit deformation.

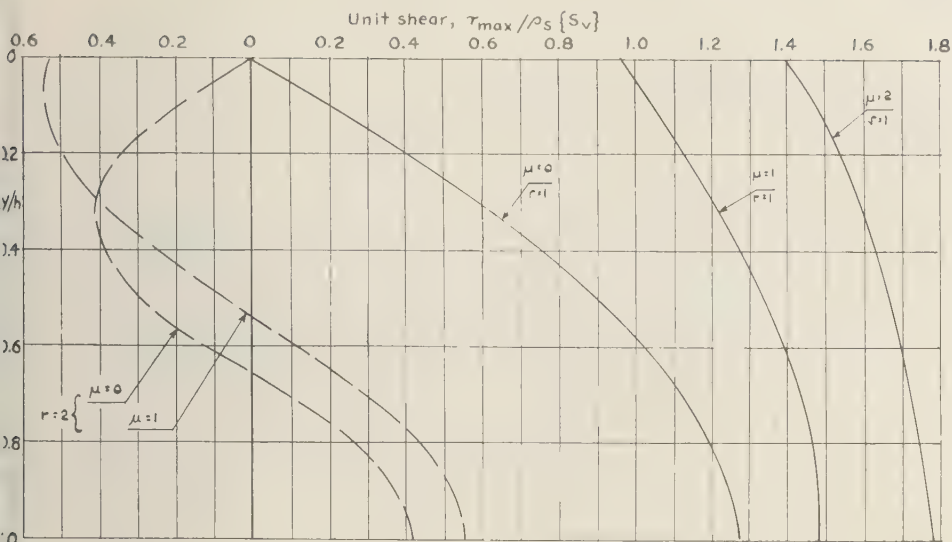


Fig. 3. Effect of load factor μ on unit shears.

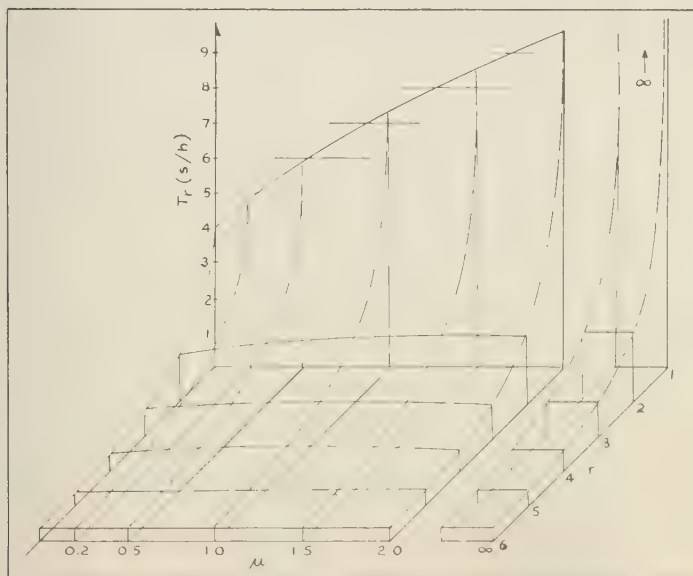


Fig. 4. Variation of natural period and higher harmonics of overburden with μ . T_r = undamped period of r -th mode; h = thickness of stratum; s = shear wave velocity.

$$\omega_r = a_r(s/h)$$

and S_d is the form of the Duhamel integral for a single degree of freedom, damped oscillator, having a natural frequency ω_r and a fraction of critical damping λ .

The equations are set up summarily because a somewhat analogous problem has been treated in greater detail elsewhere [Ambraseys, 1959].

From (4) we can easily find an expression for the values of maximum strains and stresses in the overburden in terms of the velocity spectrum of the arbitrary disturbance $g(t)$. These are given by

$$\epsilon_{\max} = s^{-1} \sum D(\mu, r) [\mu a_r \cos(a_r y/h) + \sin(a_r y/h)] \{S_s\} \quad (5)$$

$$\tau_{\max} = s^2 \rho \epsilon_{\max} \quad (6)$$

where $\{S_s\}$ is the value of the velocity of the ground motion derived from the appropriate velocity spectrum of $g_0(t)$ with the appropriate values for ω_r .

For numerical applications, values of a_r and $D(\mu, r)$ are given in Tables 1 and 2, respectively. Figures 2 and 3 show the effect of the load factor μ on the distribution of deformations and stresses within the overburden. Figure 4 shows the variation of the natural period of the overburden, expressed in a dimensionless form

(Ts/h) , with the load factor for the first six modes of oscillation.

As an application, let us consider a contact pressure of 15 psi acting on a layer of alluvial deposit 16 feet thick, the bulk density of which is 138 lb/ft³. Furthermore, let it be assumed that the velocity of the shear waves in the deposit is 100 meters/sec, and that the fraction of critical damping is 10 per cent. Assuming a base disturbance, expressible by the velocity spectrum of the December 30, 1934, earthquake (Olympia, Wash., N-S component) [Housner, 1956], acting at the base of the deposit, we can easily find that the maximum shear of 29 psi will develop at the base of the layer, whereas a shear of 14 psi will result at the top. Strains will vary between 0.5 per cent at the top of the layer and 1.0 per cent at its base, and the periods of the first two modes of free undamped oscillation will be 0.35 sec, and 0.09 sec.

REFERENCES

- Ambraseys, N. N., A note on the response of an elastic overburden of varying rigidity to an arbitrary ground motion *Bull. Seis. Soc. Am.*, 49, 211-220, 1959.
Housner, G. W., Limit design of structures to resist earthquakes, *Proc. 1st World Conf. Earthq. Eng.*, Berkeley, Calif., pp. 5-3, 5-13, 1956.

(Manuscript received August 10, 1959.)

Recurrent Geomagnetic Storms and Solar Prominences

B. BEDNÁŘOVÁ-NOVÁKOVÁ

Geophysical Institute, Czechoslovak Academy of Sciences, Prague, Czechoslovakia

In January 1959 the *Journal of Geophysical Research* published a paper by R. T. Hansen citing my paper. An explanation of the terminology and of why my results are opposite to those of Hansen is necessary. Hansen mentions the connection between quiescent prominences and geomagnetic storms, and by comparing the character figures for dark $H\alpha$ flocculi from 1917 to 1944 he shows that no such connection exists. On the other hand, from my work a certain connection between recurrent geomagnetic storms and one kind of filament-prominences, which I have called free filaments, clearly does exist. These filaments, however, are in no respects identical to quiescent prominences as they are usually defined in the literature. Moreover, they are not to be identified with all the formations included in the character figures used by Hansen.

Free filament-prominences are those that are not bound by any attraction center (they do not come from or pass into any active region). These are filaments disintegrating in the direction of the parallels in higher heliographic latitudes. Central meridian passage (C.M.P.) of the regions in which they occur often corresponds to C.M.P. of the UM regions [Bednářová-Nováková, in press] of Babcock and Babcock (1955). They are the last stage in the development of active solar regions [Bednářová-Nováková, 1953] from the diagram of Kiepenheuer [1953]. Their connection with recurrent geomagnetic storms is obvious, but it is limited by the negative effect of spots and faculae. It is the same effect that pointed out by Pecker and Roberts [1955]. It is easily determined, particularly in the period of M regions of Bartels [1932], which it disturbs, Waldmeier [1946] pointed out 13 years ago.

It is visible, however, even in periods of greater solar activity [Bednářová-Nováková, MSb]. This is how one kind of filament is connected with recurrent geomagnetic storms. Obviously, the situation is not simple. It is quite clear that:

1. Not all prominence-filaments can be taken into consideration simultaneously, since different types exhibit quite a different connection. Each type must be considered separately.

2. Attention must be paid to the position of the filament.

3. The existence of a negative effect of spots and faculae must be taken into consideration.

It is therefore apparent that a connection cannot be found by using simple statistical methods.

I am working on a paper in which I am trying to explain in greater detail all the above assertions about the hypothesis of the connection between filament-prominences and geomagnetic storms.

REFERENCES

- Babcock, H. W., and H. D. Babcock, *Astrophys. J.*, **121**, 349-366, 1955.
- Bartels, J., *Terrestrial Magnetism and Atmospheric Elec.*, **39**, 201, 1934; **37**, 1, 1932.
- Bednářová-Nováková, B., Aktivní oblasti sluneční v období kolem minima 1954 a jejich vztah ke geomagnetické aktivitě, MSa.
- Bednářová-Nováková B., MSb.
- Bednářová-Nováková, B., *Studia Geophysica et Geodetica*, no. 1, 1960, in press.
- Kiepenheuer, K. O., *Atti dei convegni del accad. naz. dei Lincei*, **11**, 107, 1953.
- Pecker, J. C., and W. C. Roberts, *J. Geophys. Research*, **60** (1), 33, 1955.
- Waldmeier, M., *Terrestrial Magnetism and Atmospheric Elec.*, **51**, 538, 1946.

(Received August 25, 1959.)

Wind Determination from an Aerobee Firing

HEINZ POETZSCHKE

*U. S. Army Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland*

In connection with the IGY program, Aerobee rockets were launched at Fort Churchill, Canada. For the purpose of obtaining wind and temperature data at high altitudes, grenades were fired from these rockets. At the firing of the Aerobee SM 105 on August 19, 1957, the positions of the grenade bursts were recorded by a photogrammetric camera system consisting of Wild BC-4 instruments. In addition, smoke puffs from the grenade bursts were illuminated by the sunlight, and the drifting trails could be identified on the photographs.

Two corresponding BC-4 photographs, taken at Metro and Twin Lakes stations with a 210 mm Infratar lens and a 115 mm Infragon lens, respectively, were oriented on a Wild-Autograph A-7. The relative orientation was

obtained with a number of stars recorded on both plates, and all the baseline components were set equal to zero. The absolute orientation of the model was then established with the help of the spatial coordinates of the grenade bursts derived from an analytical reduction, the result of which was published in BRL Technical Note No. 1206. In order to determine the spatial positions of the smoke trails, each grenade burst and the end of each corresponding smoke trail were plotted. The ends of the trails were somewhat difficult to determine and the plotted positions are correspondingly uncertain.

Table 1 presents the XYZ coordinates of the grenade bursts as extracted from the aforementioned publication, the XYZ coordinates of the endpoints of the trails as measured on the Auto-

TABLE 1

Grenade	Time from Launching <i>t</i> , sec	Burst			End of Smoke Trail		
		<i>X</i> , m	<i>Y</i> , m	<i>Z</i> , m	<i>X</i> , m	<i>Y</i> , m	<i>Z</i> , m
1	38.77	-1874	449	26481	-2187	325	26605
2	41.77	-2104	490	29930	-1878	592	30035
3	44.89	-2344	528	33371	-1810	10	33451
4	48.21	-2588	571	36932	-2905	315	37168
5	51.49	-2833	607	40316	-2565	405	40536
6	55.03	-3097	651	43842	-3463	1340	44023
7	58.65	-3366	696	47329	-4310	255	47513
8	62.30	-3637	727	50690	-3860	25	50953
9	66.17	-3925	774	54130	-3672	365	54573
10	70.18	-4223	816	57547	-5192	382	57841
11	74.50	-4542	861	61011	-4465	-435	61154
12	78.96	-4868	894	64408	-5383	-262	64588
13	83.74	-5223	943	67852	-5828	-365	68140
14	94.54	-6021	1045	74795	-5435	300	75017
15	100.48	-6465	1113	78193	-6195	1790	78747
16	107.23	-6953	1157	81526	-7855	1260	82023
17	115.19	-7539	1214	84927	-7165	1145	85226
18	124.83	-8246	1276	88250	-8055	1652	88303
Camera shutter closing	139.71						

TABLE 2

Grenade	Altitude of Burst Z , m	Distance d , m	Time Interval Δt , sec	Wind Velocity V , msec ⁻¹	Azimuth A , deg
1	26481	359	100.95	3.6	201.6
2	29930	269	97.94	2.7	24.3
3	33371	748	94.83	7.9	315.9
4	36932	471	91.50	5.1	218.9
5	40316	401	88.22	4.5	323.0
6	43842	801	84.68	9.5	118.5
7	47329	1058	81.06	13.1	205.0
8	50690	782	77.41	10.1	252.5
9	54130	654	73.54	8.9	301.7
10	57547	1102	69.53	15.8	204.1
11	61011	1306	65.22	20.0	273.4
12	64408	1278	60.75	21.0	245.0
13	67852	1470	55.97	26.3	245.2
14	74795	974	45.18	15.6	308.2
15	78193	916	39.23	23.4	68.3
16	81526	1035	32.48	31.9	173.5
17	84927	484	24.52	19.7	349.6
18	88250	425	14.88	28.6	63.1

graph, and the times as obtained from the time record of the firing.

An attempt was made to determine wind velocities from the results of the photogrammetric evaluation, whereby it was assumed that grenades were of the same performance, and that all smoke trails were cut off by the shutter closing. The spatial distance between grenade first and end of the trail was computed from

$$d = (\Delta X^2 + \Delta Y^2 + \Delta Z^2)^{1/2}$$

and the corresponding time difference, as camera shutter closing time minus burst time. Then, the velocity is obtained as

$$V = d \Delta t^{-1}$$

In addition the azimuth of each trail was computed, zero azimuth being coincident with the X axis of the coordinate system. The wind velocities and correlated parameters are compiled in Table 2.

The original plotting scale was 1:33333, and the scale of the photogrammetric model was 1:200000. The origin of the Cartesian XYZ coordinate system is at Aerobee Launcher, the minus X axis running to the south and the plus Y axis running to the east.

(Received September 14, 1959; revised November 4, 1959.)

The Profile of Solar Hydrogen-Lyman- α

J. D. PURCELL AND R. TOUSEY

*U. S. Naval Research Laboratory
Washington 25, D. C.*

The profile of the Lyman- α line of hydrogen in the radiation from the sun was determined using a photographic spectrograph, flown in an Aerobee-High rocket on July 21, 1959, as a part of the United States effort in the International Geophysical Cooperation, 1959. The instrument employed a 50-cm radius, 1200 l/mm diffraction grating in the 13th order, following a design originated by Dr. Robert J. Meltzer of the Bausch and Lomb Optical Company. Overlapping orders and stray light were effectively eliminated by means of a similar grating as a predisperser; this grating was used at an angle of incidence such that the optical system as a whole was stigmatic. Therefore, points along the spectrum line corresponded to points in a narrow slice diametrically across the sun. The dispersion was 2.6 mm/A.

High speed was achieved by coating the gratings with fresh aluminum and then magnesium fluoride as described by *Hass and Tousey* [1959], so that a reflectance of 80 per cent was attained. Another increase in speed was realized by the use of Schumann-type film produced by the Société Kodak-Pathé of Paris, and described by *Audran* [1958]. The entire instrument was aimed at the sun with a pointing control constructed by F. Wilshusen and associates of the University of Colorado.

Of ten exposures, nine were successful. Exposure times varied between 4 and 120 sec, mean altitudes between 92 and 197 km. The solar zenith angle was 69°. An exposure of 30-sec duration, made at 134- to 163-km altitude, is reproduced in Figure 1 along with a micro-photometer tracing.

The Lyman- α emission line, 1215.67 Å, is broad with wings extending about 1 Å on either side. The width at half-maximum is of the order of 1 Å. Neither the He II lines at 1215.17 and 1215.09 Å nor Lyman- α of deuterium at

1215.34 Å can be seen in the wing. The center of the line is depressed, producing two maxima separated by approximately 0.4 Å. The depression consists of two parts: a broad, weak reversal, and a deep, narrow central absorption core.

The width of the central core at half-maximum is 0.04 to 0.05 Å, and its observed central optical depth is of the order of $\frac{1}{2}$. The instrumental profile, determined before flight, was approximately 0.03 Å wide at half-maximum. Corrected widths for the absorption core, assuming Gaussian profiles, ranged from 0.02 to 0.04 Å. These values correspond to the Doppler broadening for temperatures 800° to 2100°K. Therefore this absorption core is attributed to hydrogen between the *E* region of the earth's atmosphere and the sun, but lying outside the sun's atmosphere. This temperature is consistent with most of the present atmospheric models, possibly excepting the high temperature model suggested by *Chapman* [1957].

It is believed that the central core is centered at 1215.67 Å and can be used as wavelength standard, because it is unlikely that a large enough fraction of the hydrogen would be in motion along the line of sight to produce an appreciable Doppler shift in the core as a whole.

The broad central absorption is thought to originate in the solar atmosphere. Qualitatively the profile of Lyman- α resembles Ca II, H α , and K α , and the corresponding lines of Mg II at 2795.5, 2802.7 Å. These are described, respectively, by *Goldberg and others* [1959], for example, and by *Purcell and others* [1958]. The different exposures show that the profiles of Lyman- α from plage regions, areas with no plage activity, and regions near the limb are not exactly alike. The emission line in most of the

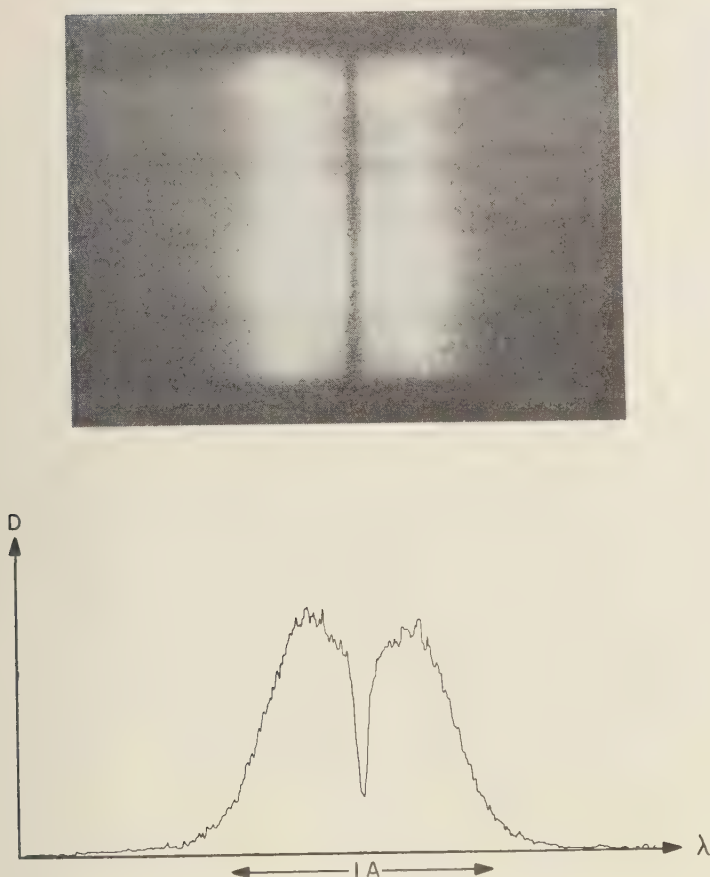


Fig. 1. The Lyman- α line in a 30-sec exposure, made between 134- and 163-km altitude, and a microphotometer tracing of the densest portion.

plage regions was shifted to short wavelengths by about 0.025 Å, corresponding to a velocity of ascent of 6 km/sec, and the maximum on the short-wavelength side was the stronger. One plage region was noted where the asymmetry was reversed, however. The optical depth of the central reversal, corresponding to H_β and K_β , reached values between 0.1 and 0.5. The sharpness of the upper and lower edges of the spectrum indicates that the pointing control performed very well, holding the image to within less than 1 min of arc.

The Lyman- α emission from the entire sun, measured during flight with an ion chamber, was 6 erg/cm² sec. From photometry of the

spectra and interpolations across the base of the absorption core it was determined that the energy removed by the narrow core was approximately 0.1 erg/cm² sec. If this energy is reradiated in 4π steradians, the sky should glow at 1215.6 Å with an intensity of the order of 0.01 erg/ster cm² sec. This is the order of magnitude of the glow reported by *Kupperian and others* [1959] from a rocket experiment in which the sky was scanned with a Lyman- α -sensitive ion chamber. Likewise this energy is compatible with their observation that the earth's albedo in daytime for Lyman- α , viewed from 120 km, is less than 2 per cent.

The hydrogen content was derived from the

observed equivalent width of the core, whose maximum value was 0.026 Å, and the f value 0.4162 for Lyman- α . The result was approximately 5×10^{13} atoms/cm² column for the slant path; this would be reduced to 2×10^{12} for a vertical column if the hydrogen is assumed to be located near the earth.

REFERENCES

Hass, G., and R. Tousey, Reflecting coatings for the extreme ultraviolet, *J. Opt. Soc. Am.*, **49**, 593, 1959.

Audran, R., *Wissenschaftliche Photographie*, **279**, Verlag Dr. Othmar Helwich, Darmstadt, 1958.
 Chapman, S., *Smithsonian contribution to Astrophys. J.*, **2**, 1, 1957.
 Goldberg, L., O. C. Mohler, and E. A. Müller, *Astrophys. J.*, **129**, 1959.
 Purcell, J. D., A. Boggess III, and R. Tousey, *IGY Rocket Rept. Ser. 1*, 198, July 30, 1958.
 Kupperian, J. E., Jr., E. T. Byram, T. A. Chubb, and H. Friedman, *Planet. Space Sci.*, **1**, 3, 1959.

(Received December 17, 1959.)

Geomagnetic and Solar Data

INTERNATIONAL DATA ON MAGNETIC DISTURBANCES

PART 1: SUDDEN COMMENCEMENTS AND SOLAR-FLARE EFFECTS FOURTH QUARTER 1958

This report continues the series which has appeared regularly in this JOURNAL since Vol. 64, No. 3, 295 (1949). Please refer to that first report for an explanation of the data given, and Vol. 59, No. 3, 423 (1954) for the definition of Ap.

Preliminary Report on Sudden Commencements
S.c.'s given by five or more stations are in *italics*. Times given are mean values obtained from normal magnetograms.

Sudden commencements followed by a magnetic storm or a period of storminess (s.s.c.)

1958 October 01d 04h 10m: AA (si: MB).—01d 15h 09m: Cm Fr.—20d 10h 21m: Tn (si: 1; bs: ks).—22d 03h 15m: sixty-one.—24d 07h 15m: forty-eight (ssc: 44; si: 3; bs: 1).—27d 15h 15m: fifty (ssc: 45; si: 5).—28d 06h 50m: thirty-nine (ssc: 30; si: 8; pt: 1).

1958 November 01d 09h 59m: fourteen (ssc: 3; si: 5; sfe: Eb El).—10d 05h 10m: Me Qu.—10d 01h 09m: forty-three (ssc: 37; si: 6).

1958 December 02d 09h 02m: AA? Ba Hr?.—02d 00h 35m: sixty-three.—13d 00h 01m: fifty-five (ssc: 54; si: 1).—13d 11h 48m: six (ssc: 5; si: 1).—15d 20h 22m: forty-six (ssc: 34; si: 12).—16d 15h 47m: forty-seven (ssc: 46; si: 1).—17d 17h 17m: forty-three (ssc: 26; si: 17).

Sudden impulses found in the magnetograms (s.i.)

1958 October 14d 04h 19m: Wn Ha AA (bs: Vl).—19d 07h 23m: Su IK AA (sfe: Lr Hr? a).—22d 16h 28m: Sw Wn Cm (bs: Vl Fr).—23d 03h 24m: AA Tn.—27d 17h 42m: eight (si: ssc: 3).—27d 18h 26m: ten (si: 5; ssc: 1; pt: 1; g: 2).—28d 11h 50m: eleven (si: 9; ssc: 1; pt: 1).—28d 13h 34m: nine (si: 5; ssc: 4).—31d 03h 17m: Cm (ssc: Tr).—31d 16h 13m: forty-three (si: 35; ssc: 4; bs: 3; sfe: Fr?).

1958 November 01d 14h 51m: fifteen (si: 10;

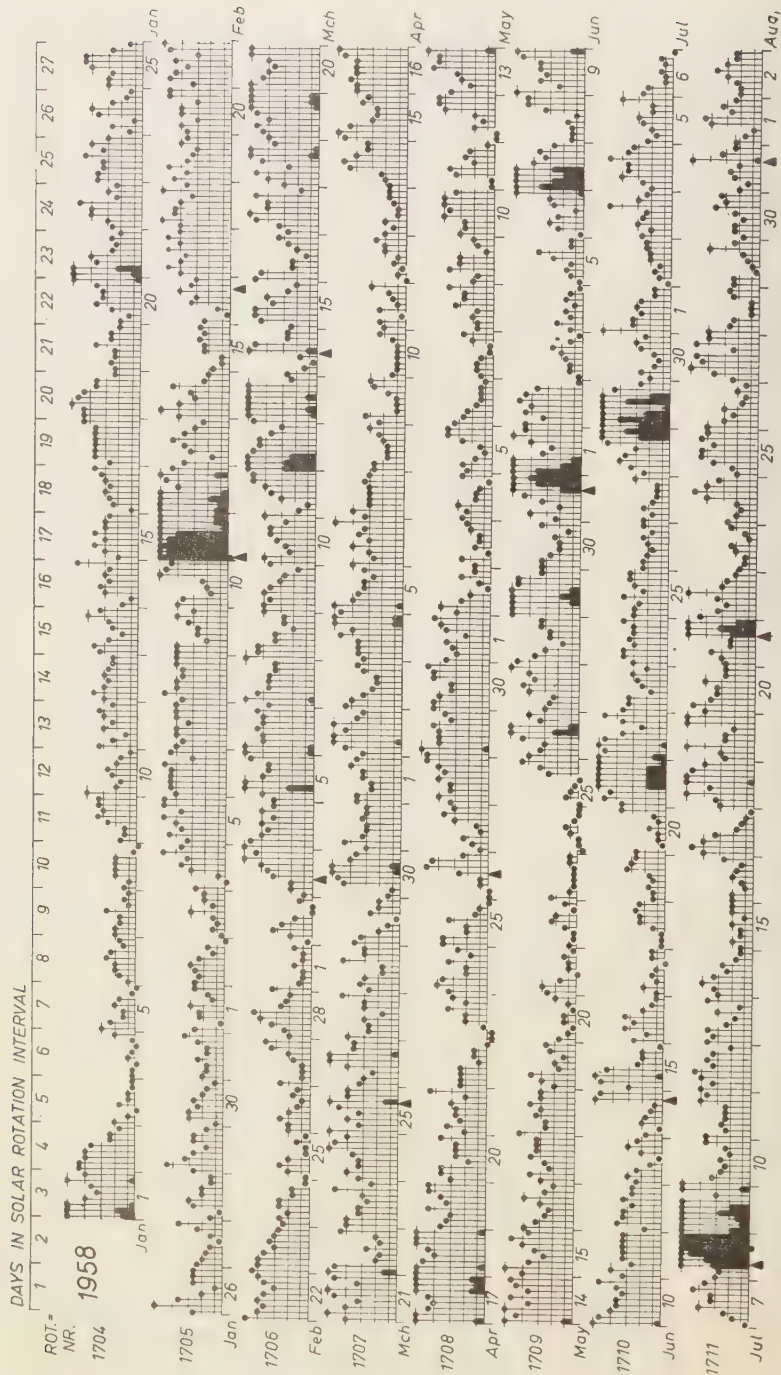
ssc: 3; pt: 1; sfe: Hu).—02d 17h 26m: Vi Va (bs: Qu).—10d 10h 55m: twenty (si: 16; ssc: 3; bs: 1).—11d 01h 29m: eighteen (si: 10; ssc: 8).—16d 02h 19m: sixteen (si: 8; ssc: 6; b: 1; sfe: Gu).—16d 04h 33m: twelve.—23d 13h 41m: Ma Fr Ba.—28d 10h 56m: MB Ba Bi.

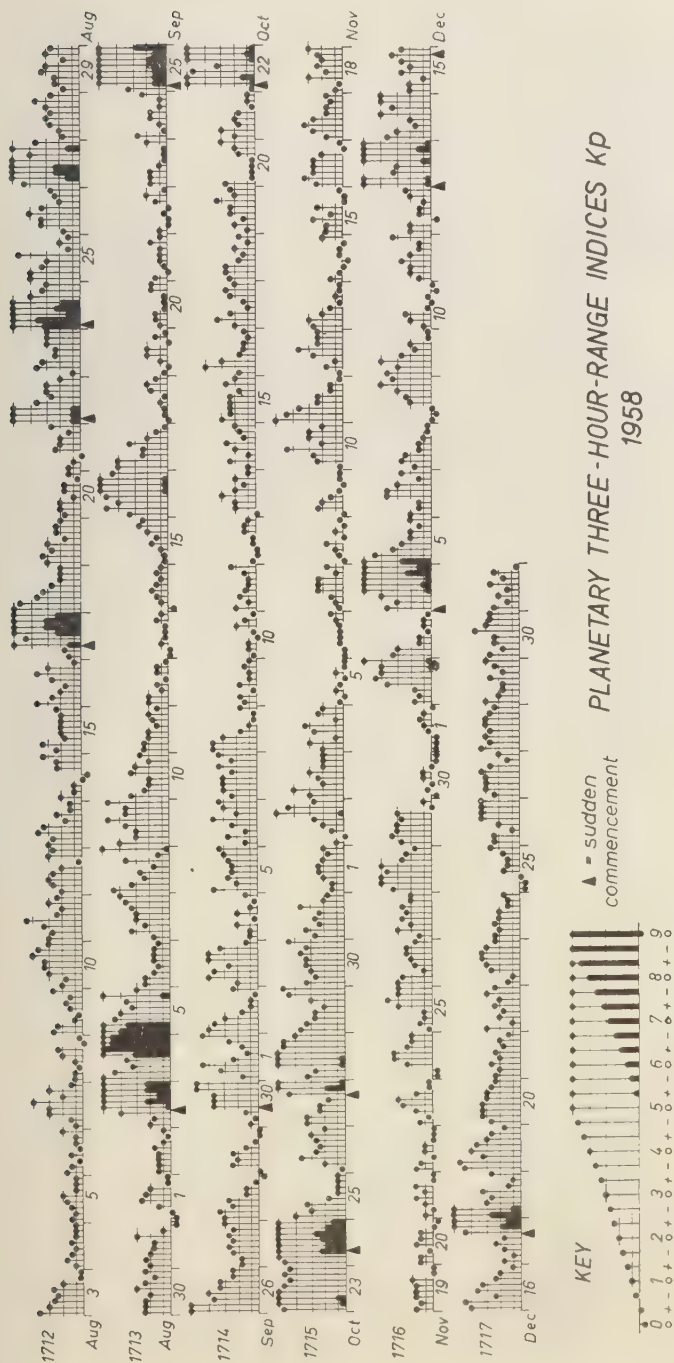
1958 December 02d 09h 44m: nine (si: 5; ssc: 3; sfe: Wn?).—04d 08h 49m: VI Ha TI MB (ssc: Le Ma; pg: Cm).—04d 15h 36m: MB Te Pi.—05d 11h 56m: six (si: 5; bs: 1).—05d 16h 09m: Cm Ma Ba.—07d 05h 29m: HI AA.—13d 12h 24m: eight (si: 7; ssc: 1).—14d 04h 54m: nine.—14d 13h 08m: thirty-two (si: 21; ssc: 7; bs: 1; sfe: Le HI Fr).—15d 09h 30m: seven (si: 5; ssc: 1; bp: 1).—16d 02h 00m: Ha Ta Tn Pi (ssc: Ks; bp: CF).—16d 11h 17m: Cm IK Qu Ta).—16d 18h 15m: Pi (ssc: Fr).—17d 23h 32m: Wn (pg: Cm).—18d 13h 07m: TI MB.—23d 12h 49m: VI CF (sfe: Fr).—25d 23h 30m: forty-six (si: 28; ssc: 14; bs: 1; bp: 1; bps: 1; sfe: Ap).—26d 01h 03m: Ky Bi El Hr.—30d 12h 50m: Cm Ma Aq Pi (ssc: SF Ba; sfe: Vl).—30d 15h 39m: eleven (si: 8; ssc: 3).

Preliminary Report on Solar-flare Effects (s.f.e.)

Effects confirmed by ionospheric or solar observations are in *italics*.

1958 October 02d 10h 14m: Wi Vl.—03d 15h 36m—15h 41m: Es.—03d 20h 42m: Si (bps: Fr).—04d 13h 53m—14h 52m: Le Es.—05d 06h 52m: El? (si: Bi).—06d 12h 38m: Te.—08d 20h 18m: HI.—11d 22h 09m: HI.—13d 11h 04m—11h 50m: Le Es Wn Wi Cm Hr?.—13d 14h 17m: Hu.—13d 19h 15m: Te Hu Pi (b: Bi).—14d 11h 30m: Le.—14d 13h 11m: HI.—14d 13h 47m: Vl.—15d 10h 23m—10h 50m: Wn Cm.—15d 13h 54m: Hu.—15d 14h 53m: Hu (si: Pi).—19d 13h 09m—13h 50m: Le Wn Wi.—20d 11h 17m—11h 21m: Es.—21d 19h 51m: Hu.—21d 23h 24m: Ap.—27d 20h 32m: Pi (si: Wn; pg: Cm; bps: AA).—31d 22h 13m: Fr.





<i>R</i>	Rot- No.	1st day	<i>C9</i>									
	1697	J24	47	45	77	67	42	11	2	1	42	363
	98	J21	63	31	213	121	263	362	242	156	221	23 35
	99	A17	23	35	1	253	666	688	836	314	212	164 333
	1700	S13	164	333	1	78	16	52	287	535	33	554 673
	01	O10	554	673	21	236	552	223	454	222	3	565 665
	02	N 6	565	665	435	531	623	123	677	64	364	326 653
	03	O 3	326	653	557	652	545	355	411	145	1	577 5
	04	D30	577	5	121	132	344	444	563	565	535	422 422
	19	J26	422	422	221	466	665	517	552	466	666	652 112
	58	F22	652	112	411	566	665	555	775	546	676	655 666
	1708	M21	655	666	443	655	656	654	221	211	566	776 441
	09	A17	776	441	243	336	665	312	321	335	246	664 453
	1710	M14	664	453	111	2	665	737	761	124	746	654 245
	11	J 7	654	245	311	227	764	542	782	414	531	417 445
	12	A 3	417	445	441	146	657	515	536	323	443	311 131
	13	A30	311	131	244	331	237	531	164	754	733	211 78
	14	S26	211	787	4	563	111	74	111	117	7	622 154
	15	O23	622	154	351	334	311	111	122	222	17	773 366
	16	N19	773	366	553	143	211	111	342	111	222	111 112
	1717	D16	111	112	322	531	152	763	133	211	753	466 542
			466	542	221	133	423	2	116	655	674	322 245

Symbol	1	2	3	4	5	6	7	8	9
<i>R</i> = 0	$\frac{1}{15}$	$\frac{16}{30}$	$\frac{31}{45}$	$\frac{46}{60}$	$\frac{61}{80}$	$\frac{81}{100}$	$\frac{101}{130}$	$\frac{131}{170}$	$\frac{171}{210}$
<i>C9</i> = 0	1	2	3	4	5	6	7	8	9
<i>Cp</i> = $\frac{0.0}{0.1}$	0.2	0.4	0.6	0.8	1.0	1.2	1.5	1.9	2.0
	0.3	0.5	0.7	0.9	1.1	1.4	1.8		2.5

Daily Indices *C9* (scale 0 to 9)

arranged in solar rotations

1958 November 03d 18h 56m: Pi.—04d 11h 59m: Hu? Pi (si: Db Ba; bs: Te).—05d 10h 15m—10h 40m: Le Es Wn Cm Lr Bi Tn (si: Ba El).—12d 16h 02m: Te.—13d 14h 20m: Te.—13d 15h 56m: Pi.—14d 00h 37m: Wa.—23d 09h 12m: El? (si: AA?; ssc: Ba).—23d 12h 50m: El?.—24d 21h 15m: Pi.

1958 December 01d 17h 37m: Hu Pi (si: Hr; ssc: Ba).—01d 18h 02m: Hu Pi.—06d 13h 13m—13h 18m: Es (bps: Am).—09d 13h 01m—13h

10m: Le Es.—09d 15h 47m: Fr.—09d 17h 00m: 17h 45m: Le.—10d 16h 19m: Hu.—11d 11 24m—11h 40m: Tn.—11d 18h 09m: Fr Pi (bs Tu).—11d 19h 35m: Fr Hu Pi.—12d 01h 06m Mb Ka Ky Ap Wa To.—12d 06h 45m: Hr.—12d 12h 57m—13h 25m: Le Es Wn Wi Vl Cm H Ma Pr Aq Eb Fr Ta Hu Lr Tn Va Pi Hr Tw (si: Db CF Fu Lg Tl SF MB Bi El; ssc: Ba).—12d 17h 59m: Hu.—18d 12h 59m: Fr.—18d 13 59m: Fr.—19d 16h 17m: Te.—21d 11h 26m—11

: Es.—24d 09h 47m–10h 12m: Le Es Cm Lr
 Hr (si: Bi El; ssc: Ba).—28d 13h 44m–14h
 : Es.—29d 10h 35m–10h 42m: Lr (pt: MB).—
 15h 34m: Hu Pi Tw.—31d 13h 22m–14h
 : Le.—31d 14h 37m: Fr Hu (si: Pi).—31d
 00m: Fr Hu (si: Pi).

*Ionospheric or solar disturbances without
 clear geomagnetic effect*

1958 October: None.

1958 November: None.

1958 December: None.

Minor disturbances reported by one station
 y are listed in the De Bilt quarterly circular,
 omitted here.

GEOMAGNETIC ACTIVITY 1958

The December 1958 data conclude the series
 the Kp-indices for the International Geophys-
 ical Year. It is a pleasure to express thanks to
 collaborating observatories for their care-
 fully derived data, namely: Sitka, Fredericksburg
 (US Coast and Geodetic Survey); Meanook,
 Invercourt (Dominion Observatory, Ottawa,
 Canada); Lerwick, Eskdalemuir (Meteorological
 Office, Edinburgh); Hartland (Royal Green-
 wich Observatory); Lovö (K. Sjöfartstyrelsen,
 Stockholm); Rude Skov (Meteorologisch Institut,
 Charlottenlund, Denmark); Witteveen (K.
 Meteorologisch Institut, De Bilt, Holland);
 Königst (Deutsches Hydrographisches Institut,
 Hamburg); Amberley (Magnetic Survey, Geo-

physics Division, Christchurch, New Zealand).
 These data have been promptly supplied, mostly
 by air-mail, so that final Kp-indices and musical
 diagrams could be edited about three weeks
 after the close of each calendar month. Unless
 delays in the transmittal were unusually long,
 tables were also issued for the first half of each
 month. This will be continued after the IGY.

On the following three pages are the 27-day
 (musical) diagram of three-hourly Kp-indices
 for 1958, and the diagram for daily characters
 C9 for June 24, 1957 through December 1958.

For explanation of Kp, Ap and Cp, please
 see IGY Annals Vol. 4, 215–236, London 1957.
 For diagrams on Kp and C9 for previous years
 (Kp 1950–1958, Cp 1937–1958), see Beiträge
 zum IGJ (abhandl. Akad. Wiss. Gottingen,
 Math.-Phys. Klasse) Heft 3, 1958; a copy of
 that paper, with text in German and English,
 may be had from the undersigned.

This work is done under the auspices of the
 International Association of Geomagnetism and
 Aeronomy, through its Committee on Character-
 ization of Magnetic Activity, and the Permanent
 Service of Geomagnetic Indices (in the Federa-
 tion of Permanent Astronomical and Geophysical
 Services), Director: J. Veldkamp, De Bilt (see
 Int. Union of Geodesy and Geophysics, Chronicle
 No. 15, Paris 1958, p. 253–258).

J. BARTELS, *Chairman*,
 IAGA COMMITTEE ON CHARACTERIZATION
 OF MAGNETIC ACTIVITY

PART 2: K_p, A_p, C_i, AND SELECTED DAYS, FOURTH QUARTER, 1958TABLE 1.—Geomagnetic Planetary Three-Hour-Range Indices *K_p*, Preliminary Magnetic Character-figures *C*, Average Amplitudes *A_p* (Unit 2_γ), and Final Selected Days, October to December, 1958

E	October 1958									November 1958								
	1	2	3	4	5	6	7	8	Sum	1	2	3	4	5	6	7	8	Sum
1	2-	3+	4+	2+	1+	2-	3-	4+	22-	1+	1+	2o	2+	2+	2o	2-	2o	15o
2	4o	4-	3o	2-	1o	1o	1-	2+	17+	1o	0+	1+	2o	3-	5+	4+	3-	20
3	2o	4o	4-	3-	4o	4o	3+	1+	25o	3o	4o	2o	2+	3o	2o	2o	2o	20
4	1o	1o	2o	1-	2o	1o	1-	1+	10-	2o	3o	1o	3-	3+	1o	2+	2o	17+
5	4-	2o	3-	2+	2+	3-	3o	3+	22o	1-	0+	1o	1-	0+	1o	0+	0+	5-
6	1+	3-	1+	2+	4-	2o	3o	3+	20-	0+	0+	1-	1-	1-	1-	1o	1+	6-
7	2+	3o	3o	3o	2+	3+	2+	3+	23-	1-	1o	1+	2+	2+	2+	2o	1-	13-
8	4-	4-	3o	2-	1o	1-	1-	2-	16o	0+	1-	1o	2-	0+	1o	0+	1-	6o
9	1+	2-	1-	2-	1o	1o	1+	1o	10-	0+	2+	2o	1+	1+	0+	1-	1-	8+
10	1-	1o	1o	0+	1-	2o	1+	1-	8-	1-	2+	3-	4+	2o	3o	3o	2o	20o
11	2o	1o	2-	1+	1o	1o	1+	2+	12-	5o	4+	4-	2+	2+	1+	1+	1-	21o
12	1-	0+	0+	1+	1o	1+	1+	1o	6+	1-	2-	1o	3-	4-	2+	2+	2+	17-
13	0+	2o	2o	3o	2o	1+	2o	1o	14-	3-	3o	2o	1+	1+	1o	1-	1-	12o
14	1o	2+	1o	2o	3-	1+	2-	2o	14o	1+	1-	0+	0o	1-	1-	0+	2o	6o
15	3o	2+	2+	2+	2+	2-	3-	2o	19-	2-	2-	3-	1+	2+	0+	0+	1-	11o
16	2-	4o	3o	2-	1+	1+	1o	1+	15+	2+	3+	3-	3-	3-	2o	0+	1-	17-
17	2o	3+	2+	1o	1+	2+	3-	2+	17o	2+	3-	3-	3-	2-	2-	1+	2-	17-
18	2+	3-	3-	2-	1+	2o	1o	2-	15+	1o	1-	3o	2-	2+	2o	2+	3o	16o
19	1o	2o	1+	2+	2+	3+	2o	2+	17-	2-	1+	2-	2-	2-	2o	0+	0+	11-
20	3-	1-	1-	1-	1-	2-	2o	3o	12o	1+	1+	1-	2-	1+	1+	0+	0o	8o
21	2o	3o	2-	2o	2-	1-	1o	1o	13o	1o	0+	1o	2-	1o	2-	0+	0+	7+
22	1-	5+	6o	5-	3+	5+	6-	5+	36+	2+	1+	0+	1o	1o	0+	0+	0+	6+
23	5-	5+	6-	5o	5-	4+	5-	4+	39-	1-	1-	2-	3-	3o	2-	1-	1-	12+
24	5-	5o	7-	6+	7-	7o	6+	6o	49-	0o	0o	2+	3+	3+	2+	2o	2o	15+
25	4o	3o	3-	2-	1o	1o	1o	1o	16o	1-	1+	1o	1o	3o	4-	3o	3o	17-
26	1-	2o	4-	3-	2+	3-	3o	2o	19o	3o	3-	1o	2-	2o	2+	2+	2-	17-
27	3+	3+	2+	2+	2-	4+	6+	5+	29o	1+	3o	2o	3-	2+	3-	2o	2o	18o
28	3+	3-	5+	5+	5o	5-	5-	4o	35o	3+	4o	4o	4-	4o	3-	2+	2-	26-
29	4-	3+	3-	3-	2o	3o	4+	5-	26+	3-	3+	3o	3o	3-	3o	0+	1o	19o
30	3+	3-	3o	3+	3+	4-	3+	4+	27o	0o	1-	0+	1o	1o	0o	0+	0o	3+
31	3-	3+	2+	3-	2+	4-	2o	2-	21-									

E	December 1958									Preliminary C. 1958			Average amplitude, A		
	1	2	3	4	5	6	7	8	Sum	Oct.	Nov.	Dec.	Oct.	Nov.	Dec.
1	0o	0o	0o	1-	1o	1+	1o	0+	4+	0.9	0.3	0.1	15	7	2
2	2-	1+	2-	4-	4+	4o	4o	5+	26o	0.7	1.1	1.3	11	16	23
3	3+	3o	1+	1+	1o	1o	1o	1+	12+	1.0	0.8	0.4	18	12	7
4	4-	3o	4o	5+	6-	5+	7-	6o	40-	0.2	0.4	1.8	5	10	54
5	7o	4+	3+	3-	2o	2o	1+	2o	25-	0.8	0.1	1.3	13	3	28
6	2o	1+	1+	4-	4-	3o	1+	3-	19o	0.6	0.1	0.8	12	3	12
7	2-	3-	2+	1+	1+	1o	1-	1-	12-	1.0	0.2	0.2	14	6	6
8	0+	0o	0+	2o	2+	3o	4o	3+	15+	0.5	0.0	0.8	10	3	10
9	4-	3o	4o	3-	2-	2o	1-	1-	18+	0.1	0.1	0.7	5	4	12
10	1o	2o	2+	1-	0+	1-	0o	0+	7+	0.1	1.0	0.0	4	12	4
11	1+	2o	2o	3-	2o	1+	1+	3+	16o	0.1	0.9	0.5	6	17	8
12	2o	2o	0o	1+	1+	1o	2o	2+	12o	0.0	0.6	0.3	3	9	6
13	5+	5+	3+	3o	6-	5o	6o	5+	39o	0.3	0.2	1.6	7	6	50
14	4-	3o	2+	2o	4o	3+	4o	4o	26-	0.4	0.1	1.2	7	3	19
15	2o	2o	1o	3-	2o	3-	3+	3-	18+	0.7	0.2	0.7	10	6	10
16	4o	4+	2+	4-	3+	2+	2o	1o	23o	0.4	0.5	1.0	9	9	16
17	2-	1o	2-	2+	2-	4o	6o	6o	25o	0.4	0.3	1.4	9	9	30
18	7o	6-	3o	3-	3+	1+	1o	3-	27-	0.2	0.5	1.3	8	8	33
19	4+	5-	3+	4o	4-	3o	3-	2+	27+	0.6	0.1	1.1	8	5	21
20	3-	3+	3+	3+	3o	3o	2+	2+	24-	0.3	0.1	0.8	6	4	15
21	2+	2o	3-	2o	2+	2o	1+	1+	16+	0.2	0.1	0.4	6	4	8
22	3o	1o	2o	2-	2+	2o	3o	2+	17+	1.5	0.1	0.6	47	3	9
23	1+	1o	3-	3o	3+	3+	3o	2+	20o	1.4	0.4	0.8	44	7	12
24	2+	2o	2+	2o	2o	1+	1+	2-	15o	1.8	0.5	0.2	89	8	7
25	0o	0o	0+	2-	1+	1+	2+	3-	9o	0.6	0.8	0.2	10	10	4
26	2+	2+	1o	2o	3+	3+	3+	3+	21o	0.8	0.5	0.9	11	8	13
27	2+	3o	3o	3o	3+	3+	1+	2o	21-	1.4	0.6	0.7	30	9	12
28	3+	3-	3o	3-	3o	3o	3-	2+	23-	1.4	1.1	0.8	37	18	14
29	3o	2o	2-	2-	3+	2-	2+	3o	18o	1.0	0.8	0.4	20	12	10
30	3-	2-	2+	2+	4-	3o	3-	3o	21o	1.0	0.1	0.7	19	2	12
31	2+	1+	2o	1+	3-	2-	2+	1-	14+	0.8		0.2	12		7

TABLE 1.—(Concluded) Final Magnetically Selected Days, October to December, 1958

Month	Five quiet days					Ten quiet days										Five disturbed days				
1958	4	9	10	11	12	4	9	10	11	12	13	14	18	20	21	22	23	24	27	28
ber	5	6	8	22	30	5	6	8	9	14	19	20	21	22	30	2	3	10	11	28
ember	1	7	10	12	25	1	3	7	10	11	12	21	24	25	31	4	5	13	17	18
ember																				

TABLE 2.—Monthly Mean Values of Ci , Cp , and Ap , October to December, 1958

Index	October 1958	November 1958	December 1958
Mean Ci	0.68	0.42	0.75
Mean Cp	0.68	0.40	0.70
Mean Ap	16	8	15

COMMITTEE ON RAPID VARIATIONS AND EARTH CURRENTS

A. ROMAÑÁ, *Chairman* Observatorio del Ebro, Tortosa, Spain

COMMITTEE ON CHARACTERIZATION OF MAGNETIC DISTURBANCES

J. BARTELS, *Chairman*
University
Göttingen, GermanyJ. VELDKAMP
Kon. Nederlandsch Meteorologisch Instituut
De Bilt, Holland

PROVISIONAL SUNSPOT-NUMBERS FOR
APRIL TO JUNE 1959

(Dependent on observations at Zurich Observatory
and its stations at Locarno and Arosa)

Day	Apr.	May	June
1	249	108	152
2	242	112	133
3	174	113	148
4	159	105	166
5	124	96	162
6	108	138	180
7	101	156	181
8	132	188	177
9	144	252	172
10	177	268	160
11	197	295	172
12	189	285	176
13	178	265	165
14	193	240	170
15	170	204	158
16	142	182	172
17	109	187	161
18	119	198	174
19	108	185	182
20	132	187	173
21	128	151	162
22	138	145	170
23	186	149	188
24	186	143	157
25	203	178	184
26	184	188	184
27	165	177	184
28	165	132	160
29	160	99	158
30	115	106	147
31		130	
Means.....	159.2	173.0	167.6
No. days.....	30	31	30

Mean for quarter: 166.7 (91 days)

M. WALDMEIER

SWISS FEDERAL OBSERVATORY
Zurich, Switzerland

FREDERICKSBURG THREE-HOUR-RANGE
INDICES K FOR APRIL TO JUNE 1959
[K9 = 500 γ ; scale-values of variometers in
 γ /mm: D = 2.7 H = 2.5; Z = 3.0]

Gr. day	April 1959			May 1959			June 1959		
	Values	K	Sum	Values	K	Sum	Values	K	Sum
1	3431	0110	13	0312	2023	13	3221	1232	10
2	0012	1122	9	1330	1123	14	3333	3343	23
3	2222	1202	13	1121	1133	13	3533	2333	23
4	3202	0111	10	2233	2235	22	4443	3244	28
5	3201	1121	11	5533	4210	23	4322	3332	27
6	2111	1112	10	1111	1100	6	2223	2223	13
7	2211	2122	13	1011	1113	9	1312	2322	16
8	3233	3234	23	3555	3334	31	2132	2343	20
9	5423	2145	26	3222	4323	21	3444	4333	28
10	4465	6543	37	2234	2223	20	3513	2223	21
11	4433	2134	24	3242	3346	27	2124	5423	23
12	3222	2113	16	6666	5743	43	1010	1232	10
13	2141	2232	17	1343	2322	20	1000	2122	8
14	4421	2123	19	1001	2211	8	1102	3213	13
15	4221	2232	18	3233	4335	26	2232	1212	15
16	2122	0212	12	5564	2223	29	2321	0133	15
17	0032	3122	13	2133	3323	20	0000	0134	8
18	2001	1211	8	3433	3235	26	2321	3232	18
19	0111	1212	9	3323	2233	21	2421	2111	14
20	0011	2111	7	3222	2233	19	3221	2132	16
21	2223	2201	14	4322	2243	22	2121	2233	16
22	0000	1000	1	3333	1343	23	3223	2234	21
23	0004	4566	25	4322	2233	21	3321	2335	22
24	4443	4333	28	4454	4356	35	5730	0034	22
25	3432	3353	26	5443	2232	25	2202	1023	12
26	3322	2223	19	4102	2122	14	3443	3210	20
27	3442	2212	20	3221	1012	12	0034	5444	24
28	2344	2123	21	0100	0112	5	4444	4243	29
29	5334	4322	26	1000	0013	5	2356	4455	34
30	2230	4343	21	2001	3221	11	4345	4343	30
31				3332	4443	26			

ROBERT E. GEBHARDT
Observer-in-Charge

FREDERICKSBURG MAGNETIC OBSERVATORY
Corbin, Virginia

PRINCIPAL MAGNETIC STORMS

(Advance knowledge of the character of the records at some observatories as regards disturbances)

Observatory (Observer-in-Charge)	Green- wich date	Storm-time		Sudden commencement			C- figure, degree of ac- tivity ⁴	Maximal activity on K-scale 0 to 9			Ranges			
		GMT of begin.	GMT of ending ¹	Type ²	Amplitudes ³			Gr. day	Gr. 3-hr. period	K- index	D	H	Z	
					D	H								Z
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Heger (J. Beers)	1959	<i>h m</i>	<i>d h</i>		<i>°</i>	<i>γ</i>	<i>γ</i>					<i>°</i>	<i>γ</i>	<i>γ</i>
	Apr. 9	18 28	10 23	s.c.*	-59	-365	-65	ms	10	3,5	7	256	1720	870
	Apr. 23	10 34	23 23	s.c.*	-10	+77	-37	ms	23	6,7	6	196	940	580
	May 11	14 51	12 22	s.c.*	-14	-126	-23	ms	12	2,3,4,6	7	197	1770	1170
	May 24	05 40	24 18	s.c.*	-40	+211	-77	ms	24	3,4,6	6	158	910	650
	Jun. 27	07 ..	28 23	ms	27	6	6			
									28	3,5	6	280	1510	780
									29	3,4	7	190	1380	890
	Jun. 29	03 ..	30 20	ms						
L. Clevén)	Apr. 9	00 30	09 10	ms	9	3,4	7	65	760	250
	Apr. 9	18 29	10 20	s.c.	+7	+100	+40	s	10	3	9	220	1550	1010
	Apr. 23	10 36	26 10	s.c.	-6	+55	+10	m	24	5	6	80	420	350
	May 8	05 00	08 15	s	8	4	8	75	1140	750
	May 11	23 00	12 22	s	12	3,6	8	150	1460	710
	May 16	01 00	16 10	ms	16	2	7	115	620	400
	May 18	04 00	18 10	ms	18	3	7	50	520	295
	May 24	05 39	25 10	s.c.*	-7	+70	+16	ms	24	3,4	7	155	600	600
	Jun. 23	17 00	24 09	ms	24	2	7	45	650	300
	Jun. 27	07 30	01 00	s	29	3,4	8			
L. Polet)	Apr. 9	18 29	10 22	s.c.*	-4	+67	-1	ms	10	3,6	7	50	350	195
	Apr. 23	10 37	24 17	s.c.*	-3	+30	-1	ms	23	7,8	6	55	205	145
	May 4	20 22	05 16	s.c.	-1	+26	0	m	4	8				
									5	1,2	5	30	135	60
	May 11	23 26	12 24	s.c.	+2	+124	-1	s	12	6	8	40	405	165
	May 15	08 00	16 11	ms	15	6,7	6	25	250	115
	May 24	05 40	25 07	s.c.*	+6	-35	+1	ms	24	8	8	35	280	125
	Jun. 11	09 09	11 18	s.c.	+1	+28	0	ms	11	5	7	15	215	70
	Jun. 27	13 48	28 07	s.c.	+5	-76	0	ms	27	5	7	30	220	95
Fredericksburg (E. Gebhardt)	Jun. 29	07 28	30 02	s.c.*	+8	-24	0	ms	29	8	6	30	130	100
	Jun. 30	08 02	30 24	s.c.*	+4	-12	0	ms	30	5	6	25	190	100
	Apr. 9	18 28	11 02	s.c.*	+8	+5	-3	ms	10	3,5	6	46	208	148
	Apr. 23	10 36	26 06	s.c.	+2	+38	-5	ms	23	7,8	6	41	191	170
	May 4	20 20	05 16	s.c.*	+11	-2	m	4	7,8	5	22	146	64
									5	1				
	May 7	21 ..	13 01	ms	12	6	7	34	396	163
	May 15	07 ..	16 12	ms	16	3	6	34	154	130
	May 24	05 40	25 12	s.c.*	+54	-9	ms	24	8	6	26	158	154
L. Viets)	Jun. 9	00 16	10 13	s.c.*	+2	+34	-4	m	10	2	5	20	87	54
	Jun. 11	09 10	11 23	s.c.*	+4	+32	-3	m	11	5	5	18	121	43
	Jun. 23	15 ..	24 09	ms	24	2	7	37	107	112
	Jun. 27	07 ..	29 03	m	27	5	5	24	145	82
	Jun. 29	07 28	30 24	s.c.	-2	+41	-5	ms	29	4	6	26	142	190
	Apr. 9	18 29	12 02	s.c.*	+4	+30	+2	ms	10	5	7	29	223	48
	Apr. 23	10 36	24 17	s.c.*	-2	+49	+2	ms	23	8	6	24	216	59
	May 4	20 21	05 16	ms	5	1	6	16	131	42
	May 7	21 52	08 17	m	8	2,3,4	5	15	89	27
L. Viets)	May 11	21 50	12 23	ms	12	1,6	7	31	285	47
	May 15	07 03	16 10	s.c.	m	15,16	8,2,3	5	12	102	36
	May 24	05 40	25 12	s.c.	-1	+49	+3	ms	24	8	6	21	171	42
	Jun. 11	09 09	11 17	s.c.*	-3	+56	+3	m	11	4	5	13	78	
	Jun. 29	07 28	30 17	s.c.	-2	+52	+3	ms	29	3,4	6	17	98	38

¹ Approximate time of ending of storm construed as the time of cessation of reasonably marked disturbance movements in the traces; more specifically, when the K-index measure diminished to 2 or less for a reasonable period.² s.c. = sudden commencement; s.c.* = small initial impulse followed by main impulse (the amplitude in this case is that of the main impulse only, neglecting the initial brief pulse); ... = gradual commencement.³ Signs of amplitudes of D and Z taken algebraically; D reckoned positive if towards the east and Z reckoned positive if vertically downwards.⁴ Storm described by three degrees of activity: m for moderate (when K-index as great as 5); ms for moderately severe (when K = 6 or 7); s for severe (when K = 8 or 9).

GEOMAGNETIC AND SOLAR DATA
 PRINCIPAL MAGNETIC STORMS—Continued

Observatory (Observer- in-Charge)	Green- wich date	Storm-time		Sudden commencement			C- figure, degree of ac- tivity ⁴	Maximal activity on K-scale 0 to 9			Ranges		
		GMT of begin.	GMT of ending ¹	Type ²	Amplitudes ³			Gr. day	Gr. 3-hr. period	K- index	D	H	
					D	H							Z
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
San Juan (M. Vazquez)	1959	<i>h m</i>	<i>d h</i>		<i>°</i>	<i>γ</i>	<i>γ</i>					<i>°</i>	<i>γ</i>
	Apr. 9	18 29	10 24	s.c.	-0.4	+24	-8	ms	10	3,4,5	6	15	170
	Apr. 23	10 38	24 17	s.c.	+3	+21	-4	ms	23	7,8	6	22	207
	May 11	23 27	12 20	s.c.	+1	+28	-10	ms	11	8	6	10	208
Honolulu (G.E.Haraden)	Apr. 9	18 28	10 21	s.c.	-2	+20	+12	ms	10	4	6	3	110
	Apr. 23	10 36	24 15	s.c.	-0.5	+30	+15	m	23	7,8	5	3	85
	May 4	20 21	05 15	s.c.	+1	+5	+6	ms	5	1	6	1	115
	May 11	23 ..	12 24	ms	11	8	6	10	160
									12	1,2,6			
									10	4			
									23	7,8			
Huancayo (A. Giesecke) (M. Casaverde)	Apr. 10	08 10	10 19	s.c.	-1	+66	+12	ms	10	5,6	7	10	369
	Apr. 23	10 36	24 17	s.c.	+2	+52	+8	ms	23	5,7,8	6	8	533
	Apr. 29	01 23	29 19	ms	29	5	6	5	211
	May 4	20 21	05 15	s.c.	0	+40	+4	m	5	1,2,5	5	8	281
	May 8	01 40	08 17	m	8	5,6	5	6	297
	May 11	23 27	12 20	s.c.	0	+45	+7	s	12	6	8	10	605
	May 24	05 40	25 05	s.c.	0	+28	+7	m	24	5,6,7,8	5	8	208
Binza (J. Bodson)	Jun. 11	09 09	11 18	s.c.	-1	+32	+7	m	11	5,6	5	6	223
	Jun. 27	06 30	28 19	ms	27	5,6,7	6	8	275
	Jun. 29	07 28	30 19	s.c.	+1	+19	+5	ms	29	5	6	9	182
	Apr. 9	18 29	10 24	s.c.	+1	+38	-2	m	10	4	...	11	248
	Apr. 23	10 37	24 16	s.c.	+2	+59	-5	m	23	8	386
	May 11	23 32	12 24	s.c.	+1	+53	-5	m	12	1	...	11	235
	May 24	05 43	25 06	s.c.	+1	+41	-3	m	24	3	...	8	122
Port Moresby (J. A. Brooks)	Jun. 11	09 11	11 16	s.c.	+2	+47	-4	m	11	4	...	7	139
	Jun. 29	07 31	30 18	s.c.	+1	+24	-3	m	29	4	...	3	151
	Apr. 9	18 29	10 20	s.c.	+1	+32	+28	m	7	237 (1)
	Apr. 23	10 36	24 11	s.c.	0	+32	+25	m	8	136
Elisabethville/ Karavia (A. Alexandre)	May 11	22 ..	12 20	m	10	250
	May 24	05 41	25 10	s.c.	...	+13	+14	m	6	159
	Jan. 25	08 39	26 17	s.c.	-1	+28	...	m	25	4,5	...	20	248
	Feb. 25	02 28	26 24	s.c.	-1	+24	...	ms	23	3,4,5	...	22	265
Apia (J. G. Keys)	Mar. 26	08 43	29 24	s.c.	-3	+44	+2	m	26	4,5	...	17	445
	May 11	23 29	12 24	s.c.	-1	+64	-2	ms	12	1,2,6	...	12	249
	Apr. 5	23 21	9 18	s.c.	0	+13	-5	m	9	2	5	7	179
	Apr. 9	18 29	12 14	s.c.	-1	+17	-7	ms	10	4	6	7	196
Apia (J. G. Keys)	Apr. 23	10 36	1 01	s.c.	0	+24	-8	m	23	7	5
									26	1	5	7	119
	May 4	20 20	5 17	s.c.	0	+7	-3	ms	5	1	6	4	164
	May 7	20 12	14 01	ms	8	2	6
									11	8	6	9	225
	May 24	05 41	26 04	s.c.	..0	+14	-7	m	24	3,8	5	5	113
	Jun. 3	23 00	8 11	m	4	8	5	7	101
	Jun. 11	09 08	12 03	s.c.	+1	+27	-10	m	11	4	5	4	73
	Jun. 29	07 28	30 23	s.c.	0	+14	-4	ms	29	4	6	5	110

PRINCIPAL MAGNETIC STORMS—Continued

Observatory (Observer-Charge)	Green- wich date	Storm-time		Sudden commencement			C- figure, degree of ac- tivity ⁴	Maximal activity on K-scale 0 to 9			Ranges			
		GMT of begin.	GMT of ending ¹	Type ²	Amplitudes ³			Gr. day	Gr. 3-hr. period	K- index	D	H	Z	
					D	H								Z
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Manus M. (an Wijk)	1959	<i>h m</i>	<i>d h</i>	SFE	'	γ	γ	..	8	4	4	'	γ	γ
	Apr. 8	09 15	bps	m	9	1	5	13	57	60
	Apr. 9	00 11	09 12	s.c.	+2	+27	+21	ms	10	5	7	26	152	139
	Apr. 9	18 28	10 21	s.c.	+2	+27	+21	ms	10	5	7	26	152	139
	Apr. 11	20 54	12 02	bps	m	11	8	5
	Apr. 23	10 36	24 17	s.c.	+1	+37	+30	ms	23	8	7	28	234	238
	Apr. 28	22 35	29 23	bps	m	28	8	5
	May 4	20 22	05 15	s.c.?	0	+7	+4	m	5	1	5	14	101	84
	May 7	21 ..	08 17	m	8	3	5	19	81	70
	May 11	23 28	13 00	s.c.	+1	+29	+20	ms	12	1,6	6	25	185	145
	May 15	07 04	16 12	s.c.?	+1	+3	+3	m	15	7	5
	May 24	05 40	24 18	s.c.	+3*	+14	+13	m	24	3,4	5	15	42	61
	May 24	20 28	25 11	s.c.?	0	+3	+2	ms	24	8	6	22	63	79
	Jun. 2	23 20	03 01	Bay	m	2	8	5
	Jun. 3	23 13	04 13	s.c.?	+1	+10	+5	m	4	4	5	21	71	36
	Jun. 23	21 ..	24 06	m	24	1,2	5	9	78	63
	Jun. 27	06 52	29 02	m	27	5,6	5
Angara M. (McGregor)	Apr. 09	18 27	10 22	s.c.	+5	+38	+30	ms	10	5	7	43	152	254
	Apr. 23	10 36	24 10	s.c.	+1	+26	+10	ms	23	7	6	21	123	76
	May 11	23 31	12 20	s.c.	+8	+21	+28	m	12	1	6	22	151	103
	Jun.	Nil Storms
	Mar. 26	08 42	29 21	s.c.*	+3	+59*	+3	ms	27	3,6	7	43	300	250
	Apr. 9	18 29	12 07	s.c.	+1	+46	0	ms	10	4,5	7	43	290	110
	Apr. 23	10 37	26 23	s.c.	+1	+41	0	m	23	7	5	23	170	41
	May 11	18 36	12 22	s.c.	-1	+3	0	ms	12	1	7	25	410	70
	May 24	05 41	25 10	s.c.*	-1	+35*	+2*	m	24	3,4,8	5	25	110	60
	Jun. 11	09 09	11 17	s.c.*	-3*	+45*	+2	m	11	4	5	10	80	35
Berkeley L. (Cullington)	Jun. 27	06 32	28 22	m	28	5	5	19	120	40
	Jun. 29	29 29	30 21	s.c.	-2	+20	0	ms	29	3,4	6	25	140	60
	Apr. 9	02 26	09 11	s.c.	+1	+13	-6	m	9	2	5	19	86	74
	Apr. 9	18 29	10 22	s.c.*	+5	+39	-24	ms	10	3,4,5	6	32	189	141
	Apr. 23	10 37	24 18	s.c.	+1	+35	-2	m	23	5,7,8
	May 11	14 53	13 12	s.c.	-1	+7	+2	ms	12	1	6	24	218	125
	May 24	05 41	25 12	s.c.*	+1	+37	-3	m	24	3	5	22	135	51
	Jun. 11	09 09	11 18	s.c.*	+2	+46	-5	m	11	4	4	8	76	11
	Jun. 29	07 29	30 23	s.c.	-2	+25	+3	m	29	3,4	5	21	140	62
	1957	Dec. 31	Jan 58 02 16	s	31	3,6,7,8 1,2,7	7	210	1746	754
Macquarie and (Turpie)	1958	Feb. 4	11 19	09 06	s.c.*	+2*	+37*	0	m	5	4,5	7	148	598
	Feb. 10	10 33	a	m	10	6	7	151	847	520
	Feb. 11	01 24	12 22	s.c.	-b	+b	+b	s	11	1,4	8	>189 (200)	>1934 (1442)	>1260 904
	Mar. 11	10 50	14 00	m	13	4,5	7

GEOMAGNETIC AND SOLAR DATA
PRINCIPAL MAGNETIC STORMS—Concluded

Observatory (Observer-in-Charge)	Greenwich date	Storm-time		Sudden commencement			C-figure, degree of activity ⁴	Maximal activity on K-scale 0 to 9			Ranges			
		GMT of begin.	GMT of ending ¹	Type ²	Amplitudes ³			Gr. day	Gr. 3-hr. period	K- index	D	H	Z	
					D	H								Z
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Macquarie Island (A. Turpie)	1958	<i>h m</i>	<i>d h</i>		<i>'</i>	<i>γ</i>	<i>γ</i>					<i>'</i>	<i>γ</i>	<i>γ</i>
	May 13	10 22	15 10	s.c.	+17	-150	+28	m	14	5	7	153	1228	7
	May 31	16 55	01 10	s.c.*	+19*	-200*	+169*	ms	31	7	8	(128)	1795	8
	Jun. 28	07 14	29 20	s.c.*	+2*	-31*	+24	m	29	4	8	176	1233	8
	Jul. 8	07 50	09 19	s.c.*	+27*	-550*	+451*	s	8	7	9	-c	2360	14
	Jul. 21	16 36	22 00	s.c.*	+22*	-75*	+63*	m	21	7	8	78	1012	7
	Aug. 17	06 23	18 13	s.c.*	-22*	+156*	-134*	ms	17	6	8	(149)	1425	7
	Aug. 24	01 41	25 15	s.c.	+21	-19	0	ms	24	2,4,5	7	166	1744	10
	Aug. 27	00 45	27 23	ms	27	3	9	>150	1990	12
	Sep. 4	14 14	05 20	s.c.	+82	-144	+35	s	4	7	8	(289)	1945	13
	Sep. 25	04 08	26 19	s.c.	-1	+37	-35	ms	25	5,6	7	142	1354	5
	Oct. 22	03 15	25 03	s.c.*	+5*	+88*	-64*	ms	23	3	7	198	1488	6
									24	3,4,7				
	Oct. 27	11 12	28 04	s.c.	+5	-37	+33	m	27	7	7	(126)	1050	6
	Dec. 4	00 37	05 15	s.c.*	-11*	+44*	-28*	s	4	6,7	8	(212)	1605	9
	Dec. 13	12 55	14 07	ms	13	5,6	7	(117)	1083	7
	Dec. 17	16 15	18 07	m	17	7	9	-c	2085	(102)

AMERICAN GEOPHYSICAL UNION

1515 Massachusetts Avenue, N.W., Washington 5, D. C.

Established by the National Research Council in 1919 for the development of the science of geophysics through scientific publication and the advancement of professional ideals.

QUALIFICATIONS FOR MEMBERSHIP

The membership of the AGU shall consist of Members, Associate Members, Student Members, and Corporation Members.

Those eligible as candidates for election to the grade of MEMBER shall be:

MEMBER (a) Persons who have made an active contribution to geophysical research through observation, publication, teaching, or administration. Definite evidence should be presented to the Membership Committee. "Publication" may include books, articles, unpublished manuscripts, inventions, or development of geophysical instruments.

(b) Persons who have made active practical application of geophysical research. It should be shown that the nominee's work has not been purely routine, but that it has tended to create new knowledge of, or to broaden or strengthen the application of, geophysical research. In general, the minimum qualifications for membership will be not less than three years of professional experience in some phase of geophysics.

(Continued on next page)

Cut along this line

APPLICATION FOR MEMBERSHIP

Please refer to qualifications on reverse side and designate below type of membership desired:

Member (\$10) ☐

Associate (\$10) ☐

Student (\$4.50) ☐
(1960)

Application forms for Corporation Membership are available upon request.

1. Surname First Name Middle Name

2. Preferred mailing address for publications

Permanent address

3. Place Month Day Year of Birth 4. Country of citizenship/naturalization

5. Nature of work and title and/or military rank; name and address of organization with which you are associated.

6. Check section or sections with which affiliation is desired.

☐ Geodesy

☐ Seismology

☐ Meteorology

☐ Geomagnetism and Aeronomy

☐ Oceanography

☐ Volcanology, Geochemistry, and Petrology

☐ Hydrology

☐ Tectonophysics

7. **EXPERIENCE** (List below, use added sheets as necessary)

Dates: From To Name and address of organization Title, duties, nature of work

8. **EDUCATION** (List below, use added sheets as necessary)

Dates: From To School Address Major Subject Degree, if any; year

(over)

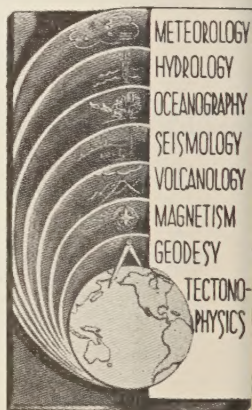
(Continued from previous page)

Those eligible as candidates for election to the grade of ASSOCIATE MEMBER shall be

ASSOCIATE MEMBER Persons who have an active interest in physical processes of the Earth or technical assistance in the application of geophysics. In general, the minimum qualification for associate membership will be acceptable training or experience in some field of geophysics or allied science.

CORPORATION MEMBER Corporations and other interested organizations shall be eligible as candidates for election to CORPORATION MEMBERSHIP. They shall have the privilege of designating a representative who has the rights and privileges of Members (use special form).

STUDENT MEMBER Those eligible as candidates for election to the grade of STUDENT MEMBER shall be persons who are graduate or undergraduate students in residence at least half-time and who are specializing in the geophysical sciences. Teaching or research assistants enrolled in more than half of a full-time academic program may also be eligible for Student Membership. Student Members shall have all the privileges of Members except that they shall not vote or hold office.



Cut along this line

*9. References: Please list below names and addresses of two or three references; include members of the AGU or others who know you well.

*10. Titles of technical contributions or publications, particularly those in the geophysical sciences, and where published.

*11. Brief statement of any special interests or qualifications in the geophysical sciences.

Date _____

Written Signature

12. (STUDENT MEMBERS ONLY) The person whose signature appears above is known to me and is a student majoring in _____ (subject) at _____

(Name of college or university) expected to graduate in _____ (year) with the degree of _____

☐ He is a full-time student, or ☐ a teaching or research assistant enrolled in more than half of full-time academic program.

(Signature of faculty sponsor)

☐ Check here if faculty sponsor is a member of AGU and willing to act as a regular sponsor for associate membership as well.

(Typed or printed name of sponsor)

(Title)

* Applicants for student membership may omit Questions 9, 10, and 11, but must fill in Question 12. Please return form with check or money order payable to American Geophysical Union, 1515 Massachusetts Ave., N.W., Washington 5, D. C.

Contents

(Continued from back cover)

Electrical Theory of Tornadoes.....	<i>Bernard Vonnegut</i>	203
The Distribution of Annual Tropical Cyclone Frequency.....	<i>H. C. S. Thom</i>	213
Airborne Measurement of Atmospheric Conductivity in Fifteen-Day-Old Thermo-nuclear Debris.....	<i>Robert V. Anderson and G. P. Serbu</i>	223
Sulfur in the Atmosphere.....	<i>C. E. Junge</i>	227
On the Dynamical Structure of the Gulf Stream as an Equivalent-Barotropic Flow	<i>G. Neumann</i>	239
On the Statistical Prediction of Ocean Temperatures	<i>Gunnar I. Roden and Gordon W. Groves</i>	249
Wave Propagation in a Medium with a Single Layer	<i>L. Knopoff, F. Gilbert, and W. L. Pilant</i>	265
Nonlinear Oscillations of a Particle in a Long Wave in a Rotating Fluid..	<i>S.-K. Kao</i>	279
A Simplified Method for the Analysis and Synthesis of Dispersed Wave Trains	<i>James N. Brune, John E. Nafe, and Jack E. Oliver</i>	287
Aerial Gravity Measurements.....	<i>Lloyd G. D. Thompson and Lucien J. B. LaCoste</i>	305
The Use of Love Waves for the Study of Earthquake Mechanism.....	<i>Keiiti Aki</i>	323
A Comparison of Rank-Difference and Product-Moment Correlation of Precipitation Data.....	<i>James E. McDonald and Christine R. Green</i>	333
Rb-Sr and K-A Ages of Rocks in Ontario and Northern Minnesota	<i>L. T. Aldrich and G. W. Wetherill</i>	337
Isotopic Evidence on the Origin and Age of the Blind River Uranium Deposits	<i>J. A. Mair, A. D. Maynes, J. E. Patchett, and R. D. Russell</i>	341
Convection Currents in the Earth's Mantle.....	<i>A. L. Licht</i>	349
Letters to the Editor:		
Soil Moisture Recounting Under a Permanent Grass Cover.....	<i>M. A. Hartman</i>	355
Criticism of the Melted Moon Theory.....	<i>Harold C. Urey</i>	358
Polar Auroral, Geomagnetic, and Ionospheric Disturbances.....	<i>E. H. Vestine</i>	360
A Note on the Effect of Surface Loading on the Shear Response of Overburdens	<i>Nicholas N. Ambraseys</i>	363
Recurrent Geomagnetic Storms and Solar Prominences..	<i>B. Bednářová-Nováková</i>	367
Wind Determination from an Aerobee Firing.....	<i>Heinz Poetzschke</i>	368
The Profile of Solar Hydrogen-Lyman-Alpha.....	<i>J. D. Purcell and R. Tousey</i>	370
Geomagnetic and Solar Data.....		373

Contents

Charge and Magnetic Field Interaction with Satellites <i>David B. Beard and Francis S. Johnson</i>	1
Satellite Orbits in an Oblate Atmosphere..... <i>D. G. Parkyn</i>	19
On the Long-Term Variation in the Cosmic Radiation..... <i>J. A. Lockwood</i>	27
Decrease of Cosmic-Ray Intensity on February 11, 1958..... <i>J. A. Lockwood</i>	39
Geomagnetic Effects on Cosmic Radiation for Observation Points Above the Earth <i>Joseph E. Kasper</i>	55
A Radial Rocket Survey of the Distant Geomagnetic Field <i>C. P. Sonett, D. L. Judge, A. R. Sims, and J. M. Kelso</i>	69
Some Characteristics of the Upper-Air Magnetic Field and Ionospheric Currents <i>Alfred J. Zmuda</i>	85
The Simultaneity of Sudden Commencements of Magnetic Storms... <i>V. L. Williams</i>	93
Geomagnetic Storm Theory..... <i>J. H. Piddington</i>	107
Occurrence Frequency of Geomagnetic Micropulsations, P_c <i>J. A. Jacobs and K. Sinno</i>	115
Daily Normals of the International Magnetic Character Figure, C_s <i>Ralph Shapiro and Frederick W. Ward, Jr.</i>	119
Lunar Daily Variation of Geomagnetic Horizontal Intensity at Alibag <i>K. S. Raja Rao</i>	123
Survey of Number of Solar Flares Observed During the International Geophysical Year..... <i>Helen W. Dodson and E. Ruth Hedeman</i>	133
The Nightly Variation of Auroras at a Subauroral Station <i>Joseph W. Chamberlain and Helene M. Thorson</i>	137
Southern-Hemisphere Observations of Sodium Emission throughout Twilight <i>B. J. O'Brien</i>	141
On the Rotation of the Polar Ionospheric Regions..... <i>C. O. Hines</i>	145
Ion-Density Measurements in the Stratosphere..... <i>John L. Kroening</i>	153
Atmospheric Tides and Ionospheric Electrodynamics..... <i>Marvin L. White</i>	173
Some Observations of Ionospheric Faraday Rotation on 106.1 Mc/s <i>R. A. Hill and R. B. Dyce</i>	177
The Lower Frequency Limits for F -Layer Radio Propagation <i>B. Fulton, O. Sandoz, and E. Warren</i>	185
A Model of the F Region Above $h_{max} F^2$ <i>J. W. Wright</i>	193
An Analysis of Time Variations in Tropospheric Refractive Index and Apparent Radio Path Length... <i>M. C. Thompson, Jr., H. B. Janes, and A. W. Kirkpatrick</i>	

(Continued inside back cover)